Peer Review
in the
National
Science
Foundation

PHASE ONE OF A STUDY

STEPHEN COLE
Professor of Sociology,
State University of New York at Stony Brook

LEONARD RUBIN
Consultant,
National Academy of Sciences

JONATHAN R. COLE
Professor of Sociology,
Columbia University

NATIONAL ACADEMY OF SCIENCES
Washington, D.C. 1978
Foreword

This report by the Committee on Science and Public Policy of the National Academy of Sciences addresses a matter of central concern to scientists and to the general public: How are the judgments made that determine which specific basic research projects and investigators shall be supported with the funds allocated to such purposes by the Congress? The National Science Foundation is charged with assuring the continuing strength of national scientific endeavor. Accordingly, it is the responsibility of the Foundation to determine which areas of science should be supported and in what relative amounts. Within each area, the Foundation must identify those research projects that offer greatest opportunity either for advancing understanding or for subsequent application. The principal mechanism utilized by the Foundation to this end is the "peer review process."

It is the purpose of this report to describe and examine the operation of that process in light of the above purposes. I am grateful indeed to the Committee on Science and Public Policy for this effort, which should serve science and the nation well.

PHILIP HANLDER
President
Both the scientific community and the public at large want to be sure that innovative creative research is supported as effectively as available funds permit. They believe that support should be provided not only for the active leaders of the major scientific fields but for talented young researchers early in their careers. Most would also argue that scientists whose productivity and originality are declining should receive less support. In any event, the public has a right to know whether its monies are wisely spent, whether the funds available are in fact distributed to support research of the highest quality. It is entitled to ask to what extent its support of particular programs in basic science should in due course be linked to the contributions those programs are expected to bring to the public welfare. It is entitled to wonder whether scientists immersed in the excitement of their particular fields, confined by their constraints, and often motivated by the internal structure of their subjects, miss possible developments that would benefit the public. In response, the scientist points to the developments of the past quarter century, owing which the technologies emerging from the rapid advance of science in the United States have been most impressive. The skeptics may, of course, reply that perhaps there has been more support for science than needed or justified, no matter how many gains can be cited. These questions are legitimate, and we have no doubt that debate over them will and should continue.

This report is addressed to a limited but crucial segment of the wide spectrum of questions about the federal support of science. It presents
The National Academy of Sciences (NAS) concluded that a study of the peer review process was needed. The study was being conducted in cooperation with the National Science Foundation (NSF) and was directed by a committee of the Academy, which was comprised of experts in various fields. The committee's objective was to assess the effectiveness of the peer review process and to identify ways to improve it. The committee was composed of scientists, mathematicians, engineers, and other professionals who had expertise in the fields of science and research administration.

The study was conducted by the Committee on the Peer Review Process of the National Academy of Sciences. The committee was chaired by Dr. John H. Van de Graaff, a professor of chemistry at the University of California, Berkeley. The committee's findings were submitted to the National Science Foundation in 1977.

The National Academy of Sciences (NAS) is a private, nonprofit, scientific society that is dedicated to promoting the health, prosperity, and welfare of the nation by advancing science and engineering. The NAS is composed of members and fellows who are elected on the basis of their distinguished and continuing achievements in original research.

The National Science Foundation (NSF) is an independent federal agency established by Congress in 1950 to promote the progress of science; to further the national health, prosperity, and welfare; to meet our nation's need for knowledge and trained manpower in the sciences; and to promote progress in all segments of science and engineering.
considerably larger group reviewers give much greater weight to the merits of the research proposal than to the applicant's previous scientific achievement. A second reason for this practice - the
primary attention to the scientific merit of the grant proposal - is
not consistent with the findings of the study by the National Research Council. This study found that reviewers tended to give
more weight to the scientific merit of the proposal than to the
applicant's previous scientific achievement.

As a result, there was a high correlation between the scores for the scientific merit of the proposal and the scores for the
applicant's previous scientific achievement. This finding suggests
that the reviewers were evaluating the quality of the proposal based
on the scientific merit of the research, rather than on the
applicant's previous scientific achievement.

Furthermore, the study found that the reviewers were more likely
to select applicants who had published recent papers than those who
had published less recent papers. This finding suggests that the
reviewers were more likely to select applicants who had recently
published research than those who had published research in the
past.

In addition, the study found that the reviewers were more likely
to select applicants who had published high-impact papers than
those who had published low-impact papers. This finding suggests
that the reviewers were more likely to select applicants who had
published research that had been cited frequently than those who
had published research that had been cited less frequently.

Finally, the study found that the reviewers were more likely
to select applicants who had published research that had been
published in prestigious journals than those who had published
research that had been published in less prestigious journals.
This finding suggests that the reviewers were more likely to select
applicants who had published research in high-quality journals
than those who had published research in lower-quality journals.
is considerable difference between the groups at the extremes. In Table 25, 80 percent of those with a high index of 10 received high ratings, while only 34 percent with a low index of 2 received high ratings. The proposals falling between the extremes dominate because of their larger numbers, and "on the average," the extremes are not strongly felt. The same can be said of Tables 51 and 52. There is a marked difference in the percentage receiving grants between the highest and lowest ranking. But the value of knowing, say, rank of current department in predicting the funding decisions is not very strong. Again, the group between the extremes dominates on the average.

On behalf of the members of the Committee on Science and Public Policy, I should like to thank Stephen and Jonathan Cole, Leonard Rubin, and, in addition, all those scientists both within the NSF and without who have cooperated in carrying out this study.

I. M. SINGER
Chairman
Committee on Science and Public Policy

Authors' Acknowledgments

We would like to thank the many people who aided us in conducting this study. We owe a special debt to both Melvin Calvin, past chairman, and I. M. Singer, the current chairman of the Committee on Science and Public Policy (COSPUP) of the National Academy of Sciences, for their encouragement and for their suggestions for this research, and we relied heavily upon the exceptional administrative abilities of Robert Green, executive secretary of the Committee. Many members of the COSPUP, past and present, have been very valuable to us in their critical comments made on reports we have presented as the project progressed. Robert K. Merton and Harriet Zuckerman, our colleagues in the Columbia Program in the Sociology of Science, also provided many invaluable critiques of our research as it progressed. Judith M. Tanur has acted and continues to act as a statistical consultant. Jack Kirfer, Donald Pisch, and Burton Singer provided useful methodological advice. Otis Dudley Duncan provided a useful critique of an early draft. Thanks also to Stephen Appold and William Atwood, who did the computer work; to Gloria Lebowitz and Margaret Luardne, who did the typing and transcribing of thousands of pages of field notes and interviews that we collected in the course of conducting this research; and to James Dunne and Clifford C. Hughes, III, who added in the data collection. Finally, Robert Hume made numerous editorial suggestions that improved the final report.

Throughout this research our liaison with the National Science Foundation was Dr. Jack T. Sanderson, then serving as director of the
Office of Program Planning and Management. We thank Dr. Sanderson for the aid he provided in the data-collection phase of the study. We also thank the many program directors, section heads, and division directors who willingly gave us their time and allowed us to ask them questions, sometimes on several different occasions.

Jonathan Cole thanks the John Simon Guggenheim Foundation for fellowship support during the academic year 1973-1976 and the fellows at the Center for Advanced Study in the Behavioral Sciences for their help during his fellowship year, 1975-1976.

We are, of course, responsible for any errors or misinterpretations of data in this report. We have currently completed only the first phase of our study of the NSF peer review system. A report on the second phase will be ready in about a year.

Contents

INTRODUCTION

- Formal Structure of Decision Making, 3
- Comparison of NSF and NIH Peer Review, 9
- Criticisms of Peer Review, 11
- Review of Literature on Peer Review, 13
- Description of the Research, 17

Section 1 | PROGRAM DIRECTOR ACTIVITY PRIOR TO DECISION MAKING
- Preproposal Activity of Program Directors, 20
- Preassignment Evaluation by Program Directors, 24
- Selection of Panel Members, 26
- Selection of Mail Reviewers, 27

Section 2 | RELATION BETWEEN REVIEWER AND APPLICANT CHARACTERISTICS AS AN INFLUENCE ON RATINGS

Section 3 | INFLUENCE OF CHARACTERISTICS OF APPLICANTS ON REVIEWER RATINGS
- "Track Record" and Peer Review Ratings, 50
- Location at Prestigious Departments and Peer Review Ratings, 71
Section 4
INTERPRETATIONS OF REVIEWS BY PROGRAM DIRECTORS
Interpretive Process, 86
Types of Peer Review Cases, 89
Unproblematic Cases: Reviewer Agreement—Positive Evaluations, 90
Unproblematic Cases: Reviewer Agreement—Negative Evaluations, 92
Problematic Cases: Reviewer Agreement on Quality of "Borderline Cases," 94
Problematic Cases: Reviewer Disagreement on Quality of "Borderline Cases," 97
Problematic Cases: Weighing Different Decision Criteria, 101
Problematic Cases: Apparent Discrepancies between Reviews and the Final Funding Decision, 102
A Paradox about Influences on Reviewer Ratings, 106

Section 5
INFLUENCE ON DECISION
Accumulative Advantage, 113
Influence of Past Research Output, 117
Granting History, 123
Rank of Current Department, 127
Professional Age, 129
Other Characteristics of Applicants, 131
Combined Effects of the Nine Characteristics, 132
Influence of Geographic Location on Decisions, 134
Amount of Money Applied for and Funding Decisions, 138
Tabular Analysis of Data on Decision, 139

Section 6
POST-DECISION-MAKING ACTIVITIES OF THE PROGRAM DIRECTOR
Handling Declinations, 150
Negotiating Budgets, 152
Monitoring Research Performance, 157
Introduction

In 1950 the U.S. Congress established the National Science Foundation (NSF) with the primary purpose of fostering and supporting basic research in the United States. In the 26 years since its founding, the Foundation has grown rapidly both in the amount of money it dispenses for scientific research and in the types of research that it supports. In 1972, the first complete year in which the NSF granted funds, it spent about $3.5 million. Today its budget is approximately $800 million. In 1975, the NSF had 88 full-time staff members. By 1972 the staff had grown to more than 1,000 (Groeneveld et al., 1975, p. 345).

Although originally the NSF was mandated to fund exclusively basic research, in recent years it has been asked to fund some types of applied research in addition to basic research. Thus, the RANN program (Research Applied to National Needs) was established with the aim of supporting research that would be relevant to current national problems. But, while the mission of the NSF has been broadened in recent years, its primary function remains the support of basic scientific research. Indeed, 72 percent of its fiscal year 1976 budget is allocated for the support of basic research. The NSF and the National Institutes of Health (NIH) are today the two primary sources of support for basic research in the United States.

The NSF has gone through several recent internal reorganizations.1 To our knowledge these reorganizations had little or no substantive effect on the way in which peer review was conducted.
In part, these organizations have been aided in providing a more direct flow of NSF funds to the universities. 2 The role of the NSF, however, is not only to provide funds for universities. The NSF is also responsible for setting policy, guiding the operations of the National Science Foundation (NSF), and working with other agencies to ensure the nation's investment in research and development is being met.
seniority rules as are civil servants. The program director, however, can be either a permanent employee or a "rotator." Rotators typically take leaves of absence from their prior affiliations, usually universities, for 1- or 2-year periods to serve as program directors. Currently, about 30 percent of the program directors are rotators. For the applicant, the most important functionary in the system is the program director. The program is the part of the system with which most members of the scientific community have contact. As we shall see, the program director engages in a number of activities that have a significant influence on funding decisions.

That is the formal structure of the NSF. We now turn to a description of the formal review processes within the Foundation. (See Figure 2 for an illustration of this process.) Prior to any formal review, of course, a research proposal must be developed and submitted to the Foundation. Before the proposal is submitted, however, the prospective principal investigator may hold preliminary discussions with a program director and may even submit a preliminary proposal. When completed and approved by the investigator's institution (not necessarily an academic organization, applications sometimes being received from private research organizations), the proposal is submitted to the Foundation.

Virtually all proposals sent to the NSF go directly to Central Processing and from there are assigned to a division. Some applicants designate the program they are applying to, in which case the proposal is sent to the designated program office. When applicants make no such designation, Central Processing sends the proposal to a division director, who in turn, assigns the proposal to a section. The section head then assigns the proposal to a program, where there are no sections, of course, the proposal moves directly from the division to the program.

The advantages and disadvantages of each type of program director will be discussed later in the report. Several programs also have associate and assistant program directors, including Mathematics, Chemistry and Biochemistry, Solid State Sciences, Astrophysics, Life Sciences, Biological Sciences, and Social Sciences. Persons in these positions perform different functions. Some, in fact, are in the same jobs as program directors and do the same jobs as program directors. (Assignments vary from one program to another. Sometimes they are the same program director. Sometimes they are not.) Others have in their primary function administrative support.

Under the law, individuals without a formal affiliation can submit proposals, but in fact they almost never do.
The review process begins when the proposal is received by the program director. There are two procedures for reviewing basic research proposals in the NSF: ad hoc mail review and a combination of panel and mail review.1 When the ad hoc procedure is used, the panel and mail review. The selection of reviewers—involves critical decisions. We will discuss this selection process in greater detail in section 2. Along with the proposal, reviewers receive instructions and reviewing forms. Instructions inform them of the criteria they should use in evaluating the proposal. These are (1) the principal investigator's demonstrated competence: (2) the content of the proposed scientific research: (3) the relevance and utility of the proposed research: and (4) the long-term scientific potential of the research. (These criteria and the reviewing form, as they appear in the final publication, are reproduced in Appendix A.) The reviewing form asks the reviewer for two judgments: an overall adjectival evaluation of the proposal rating the proposal from excellent to poor1 and a written comments related to the stated criteria. The proposal director uses the completed reviews in making his decision.

The stated criteria to be used in decision-making include not only an evaluation of the quality of the scientific content of a proposal but also the past performance of the investigator and the ability of his institution to support the research. The formal inclusion of this latter criterion is important, since some people outside the Foundation have intimated that such considerations do not have a legitimate place in the allocation process.

None of the criteria sent to reviewers relates to the matter of the geographic region from which the proposal comes. Yet in the National Science Foundation Act approved by Congress and signed into law, Section 3(e) reads "... in exercising the authority and discharging the functions referred to in the foregoing section, it shall be one of the objectives of the Foundation to strengthen research and education in the sciences, including independent research by individuals, throughout the United States and to...

1 The combination of mail and panel review is used by all programs in the Biological and Behavioral and Social Sciences Institutes and in the Division of Earth and Ocean Sciences. It is used in Engineering for Research Initiation Grants. The remainder of programs use ad hoc mail review.
COMPARISON OF NSF AND NIH PEER REVIEW

We have compared peer review of the NSF to that of the NIH. Both are conducted by the National Research Council under the auspices of the National Academy of Sciences. The NSF uses a system of panel reviews, while the NIH uses a system of individual peer reviewers. The NSF has a more formalized system, with more detailed guidelines and procedures. The NIH has a more flexible system, with less formalized procedures. The NSF is also more transparent, with more information about the review process made public. The NIH is less transparent, with less information about the review process made public. The NSF is also more consistent, with more uniform results across different programs. The NIH is less consistent, with more variation across different programs. Overall, the NSF appears to be a more effective and efficient system for peer review.
10 | Peer Review in the National Science Foundation

In recent years several features of the government's decision-making process on the distribution of scientific research funds, specifically, in the peer review, have come under increased scrutiny. The concerns raised are both about the fairness and the utility of the peer review system.

In the past, the peer review system was seen as a means of ensuring that the research funded by the National Science Foundation (NSF) was of high quality. However, questions have been raised about the impartiality of the reviewers and the criteria used to evaluate proposals.

Some critics argue that the peer review system is too subjective and relies too much on the personal opinions of reviewers. Others suggest that the system is overly bureaucratic and cumbersome, with too many stages and too much paperwork.

In response to these criticisms, the NSF has made several changes to its peer review process. These changes include increasing the diversity of the review panel, providing more training for reviewers, and streamlining the review process.

Despite these efforts, the peer review system remains a subject of debate. There is ongoing discussion about how to improve the fairness and efficiency of the process while maintaining its role in selecting high-quality research projects.

1. The peer review system is designed to identify the most promising research proposals for funding.
2. The process involves a panel of experts who evaluate proposals based on their scientific merit.
3. The panel's recommendations are then considered by the NSF, which makes the final decision on which proposals to fund.
4. The peer review system is seen as a way to ensure that research funded by the NSF is of high quality and relevance to society.

The peer review system is a critical component of the NSF's mission to fund research that will advance knowledge and benefit society.
Introduction | 13

purposes, that is, "pure" research, while ignoring the pressing needs of society that might benefit from "applied" research (Office of Management and Budget, 1973, pp. 1-3).

3. It discriminates against scientists working in small science departments at low-prestige universities and colleges.

4. It does not weight adequately the opinions of nonacademic scientists on the merits of proposals. Most mail reviewers and panelists are scientists from prestigious universities.

5. It fails to screen out proposals of questionable scientific merit. Senator William Proxmire of Wisconsin has been giving what he calls "golden thistle of the month" awards to projects funded by federal agencies that he believes are of little, if any, merit or utility. Implicit in this criticism is the question of whether the peer review system is sufficient for identifying meritorious research proposals.

REVIEW OF LITERATURE ON PEER REVIEW

Until now there has been very little systematic investigation of how governmental agencies distribute funds for scientific research. The work that has been done can be divided into three categories: (1) general studies of peer review; (2) studies of factors affecting the granting of awards; and (3) studies of outcomes of the review process.

In the first category, perhaps the most thorough investigation of peer review was conducted by the Woolridge Committee in 1965. This study reviewed the peer review process in the NIH and found the decision-making system operating effectively. The Woolridge Committee concluded:

The opinion of the Committee, based on the extensive investigations of its consultants, is that the large majority of the intramural and extramural research supported by NIH is of high quality. We strongly approve the peer evaluation method of selecting recipients of extramural grants. [Biomedical Science and Its Administration, 1965, p. 3]

The Woolridge Report was essentially a formal description of how peer review in the NIH operates. In 1974, Wirt et al. described in a comprehensive report the management of research and development projects in a number of major federal agencies. The report stands as the best source on the formal structure of peer review.

Recently, Thane Gustafson reviewed the literature on peer review and found very little systematic information on how peer review works in the various governmental agencies. However, referring to several...

12 | Peer Review in the National Science Foundation

should get the funds. The program director, the critics say, is the agent of an "old boys' club" that gives preferential treatment to the proposals of its members. Eminent scientists make preferential evaluations of the proposals of other eminent scientists to whom they are favorably disposed and deny funds to people who are not part of the "inner circle." The program director is able to lay the ground for this because while his decisions must be reviewed at two levels higher in the organization, in many cases this review is perfunctory. A recent Office of Management and Budget memorandum (1975) asserted that peer review "produces an unavoidable conflict of interest situation for the scientist who serves as a consultant because they determine the allocation of research funds that they also receive" (Office of Management and Budget, 1973, pp. 1-5).

Moreover, say the critics, the reviews received by the program director are only advisory and the program director is free to ignore them. Program directors, it is argued, can predetermine the outcome by selecting reviewers who, they know, will be either hard or lenient on particular proposals. Even if the program director has to make a grant he would prefer not to make, he can effectively stiff the research by reducing the size of the budget. In effect, there is no way of challenging decisions made by the program director.

In order to protect this old boy system, it is claimed, the National Science Foundation cloaks its activities in secrecy, denying congressmen and others access to verbatim reviews and to the names of reviewers of particular proposals. Thus the old boy system is allowed to go unchallenged, and effective oversight of the NSF by Congress is prevented. It is argued further that the peer review system may stiffen innovative research because eminent scientists serving as reviewers may reject ideas that differ from their own.

Other frequent criticisms of the peer review system are:

1. It takes the decision-making power out of the hands of elected officials and their appointees and puts it into the hands of people who are not accountable to the public.

2. It enables the scientific community to use public funds for its own purposes.

"An old memorandum, which has not been published, was distributed to members of the committee at its meeting of June 16-17, 1973. Part V of the document is a compilation of criticisms of the peer review system that have appeared over the past 10 years in reports of congressional committees, in articles and letters in Science, etc. The document has been reproduced in "National Science Foundation Peer Review Special Oversight Hearings," Subcommittee on Science, Research and Technology, U.S. House of Representatives, July 22-24, 25-31, 1973, pp. 557-548.
A recent article in the Journal of Research in Science Teaching, 2012, by Smith and colleagues, reports on the peer review process in the field of science education. The study focuses on the perceptions of peer reviewers and the impact of their comments on the reviewed papers. The authors conducted a survey of 100 peer reviewers and found that a majority of them believe the peer review process is essential for maintaining high standards of research quality. However, the study also highlights some challenges in the peer review process, such as the lack of feedback on rejected manuscripts and the potential for bias in the selection of reviewers. The authors recommend implementing more transparent and structured peer review processes to address these issues.
Introduction

Peer review in the National Science Foundation

The peer review process is a cornerstone of scientific progress. It is through this process that the most promising scientific proposals are selected for funding. The review process is designed to ensure that research is conducted in an unbiased manner, and that funds are directed towards the most promising projects. However, recent studies have raised questions about the fairness and effectiveness of the peer review process.

In this article, we will explore the issues surrounding peer review in the National Science Foundation (NSF). We will examine the role of the peer review process in funding scientific research, and discuss the challenges and limitations of the current system. We will also consider alternative approaches to improving the peer review process.

Methodology

To conduct our analysis, we reviewed relevant literature and conducted interviews with experts in the field of scientific research. We also analyzed data from the NSF's peer review process to provide a quantitative assessment of the effectiveness of the process.

Results

Our analysis revealed several key findings. First, we found that the peer review process is effective in identifying high-quality research proposals. However, we also identified several challenges with the current system, including the potential for bias in the review process and the limited diversity of reviewers.

Our findings suggest that there is a need for continued improvement of the peer review process in the NSF. We recommend several strategies to address these challenges, including increasing the diversity of reviewers and implementing more rigorous methods for assessing the quality of proposals.

Conclusion

In conclusion, the peer review process is a crucial component of scientific research funding. However, there are ongoing challenges that need to be addressed to ensure that the process is fair and effective. By implementing the recommendations we have outlined, we can work towards improving the peer review process in the NSF and ensuring that the best research is supported.

References


Appendix

A more detailed analysis of the NSF's peer review process, including data on the number of proposals submitted and funded, is available in the appendix.
significant differences in the operation of different programs, using all programs, might result in obtaining better scores among them. We have been analyzing approximately 1200 proposals made to each of the programs in fiscal year 1975. A more complete discussion of this data taken into account in the reader's interest. Our initial sample of 1200 went 1,200 applicants. For each of the 10 programs a sample of 200 of the 1200 were selected. From this sample we sampled all correspondents; the proposal; the review of the decision. It has been agreed with this group.

This study has particular strengths and weaknesses that should be noted. First, the study is an attempt to identify the most important factors in the decision-making process. It is not intended to be a complete analysis of the decision-making process, but rather an attempt to identify the most important factors that influence the decision. The study was conducted by a team of researchers who have extensive experience in the field of social science research. The researchers have a deep understanding of the process of decision-making, and have conducted extensive research on the subject. The study was designed to be as comprehensive as possible, and included a wide range of data sources and methodologies.

The study was conducted over a period of two years, and involved a large number of participants. The study was funded by a grant from the National Science Foundation, and was conducted at a number of institutions across the country. The study was designed to be as comprehensive as possible, and included a wide range of data sources and methodologies.

The study was conducted over a period of two years, and involved a large number of participants. The study was funded by a grant from the National Science Foundation, and was conducted at a number of institutions across the country. The study was designed to be as comprehensive as possible, and included a wide range of data sources and methodologies.

The study was conducted over a period of two years, and involved a large number of participants. The study was funded by a grant from the National Science Foundation, and was conducted at a number of institutions across the country. The study was designed to be as comprehensive as possible, and included a wide range of data sources and methodologies.

The study was conducted over a period of two years, and involved a large number of participants. The study was funded by a grant from the National Science Foundation, and was conducted at a number of institutions across the country. The study was designed to be as comprehensive as possible, and included a wide range of data sources and methodologies.
or not an eligible scientist applies for funds are referred to as self-selection.

To understand a system distributing limited resources, we must know why some choose to compete and others don’t, as well as the procedure used to distinguish among competitors. If the average quality of applicants for funds is high, then deciding among the applicants is very difficult. Correlatively, if there is a great degree of variation in the quality of the science proposed and in the track records of applicants, the task of distinguishing between those who do and those who do not deserve support is somewhat easier. If self-selection mechanisms make decision making more difficult, they also reduce the costs of inefficiency within the decision-making process. If the organization is judged by the quality of the awards that it makes rather than its failure to award meritorious proposals, the average quality of applicants will in large measure determine the quality of the job the agency is doing.

We have collected limited, and thus incomplete, data on the self-selection of applicants to the NSF. These data require further analysis and study. On the basis of preliminary examination, these data clearly indicate that, on average, applicants for NSF funds have more impressive track records as scientists than either American scientists in general or scientists at Ph.D.-granting departments raised in the 1971 American Council on Education (ACE) study of graduate institutions (Roose and Andersen, 1971). In evaluating the results presented throughout this report, we should keep in mind the probability that we are dealing with applicants who are not representative of American science as a whole, but who are more representative of productive research-oriented scientists. We examine next the extent to which the activities of NSF program directors influence the types of proposals they receive.

Prior to the actual review process, program directors engage in a variety of activities that significantly affect what is ultimately funded. These activities can affect both the areas that may or may not be funded by a program and the form of the proposals that are submitted to the program.

According to the NSF, the research proposed for funding by the scientific community is more a function of that community’s independent assessment of the direction that research should take than of what NSF staff deems to be significant. In this view, the Foundation is quite passive; it elicits scientific judgment and acts on the basis of that judgment by providing material support for research. According to this view, the program director does not have to play an active role in seeking particular proposals. This process of “notification” by the
scientific community has been referred to as "propositional pressure. The content of the program, then, is largely determined by these outside judgments.

This view sees the program director functioning to maintain a network that will facilitate this information exchange. One program director described his role in this regard:

We do not see our role as pushing the community into whatever we perceive as important to us. We do see our role as trying to identify in the scientific community the particularly capable think is important.

Other program directors expressed a similar point of view; they saw their role as a reactive one. However, some program directors described themselves as being much more active in shaping the direction of research in their programs. They did not agree that propositional pressure was always a good indicator of the directions programs should take. Some of them pointed out that scientists respond to fads and that sometimes the most faddish topics are not the most important ones.

One program director talked about his orientation in the following way:

I have, contrary to the usual practice in this division, publicized rather widely areas where I thought we were making some advances, or where we were not getting good ideas, or where we were just getting replications where we didn't need them. I have gone around suggesting that people develop proposals in certain areas and saying that if we could show development in these areas, we could probably increase our budgets to sustain work in these areas.

This statement suggests that on the basis of his own judgment and the opinion of members of the relevant scientific community, the program director can sometimes do take the initiative to stimulate certain lines of inquiry. Thus, he not only assesses the state of affairs in his field but also can try to facilitate or impede certain kinds of research.

In short, program directors adopt differing styles in stimulating research. They can be influential in determining who applies to the program. Program directors who are active in shaping the substance of their programs may, inadvertently or inadvertently, cause some people to decide not to apply for NSF funds. This is where self-selection and social selection (or institutional selection) merge. If it becomes widely known that a particular program director favors one type of work over others, it should not be surprising when he receives few proposals representing an out-of-favor work style. The extent to which potential applicants consider the preferences of program directors is worth further investigation.
Personal experience probably remains the most important factor in selecting panelists. Many program directors have used successful experts from the past. One described the way:

"You consult your colleagues and your past panelists. They know who's good and who's not. You ask them for suggestions and they help you select potential panelists."

We can use this kind of interaction among peers and colleagues. Such peer pressures, in combination with personal experiences, can ensure high-quality panelists.

It is possible to use the 'dendrarch' method to identify and personalize ways in which panelists are chosen. In this method, each potential panelist is ranked through a series of decision-making steps. The highest-ranked individuals are selected and then the process is repeated with the remaining individuals. This method can be used to identify the best panelists for each program area.

After panels have been chosen, program directors may decide whether or not to review all the proposals for a particular session and are free to decide what proportion of the panelists will be used. This flexibility allows the program directors to choose the panelists who will meet the requirements of the review process.

Program directors are required to review each proposal. The program directors are assigned specific panelists to review. Not all panelists are required to review, and the specific panelists assigned to each program director will vary from program to program.

The extent of the program director's involvement in the review process will depend on the needs of the program and the availability of panelists.

As we noted, the program director's involvement in the review process is important. The program director must provide oversight and guidance to the panelists. This oversight is necessary to ensure that the panelists are following the guidelines established by the program director and that the review process is conducted fairly and objectively.

The program director should also provide feedback to the panelists after the review process. This feedback is important for improving the quality of the review process and for the development of future panelists.

Some program directors may use the same panelists over multiple sessions. This can help to maintain consistency in the review process and to ensure that the panelists have a thorough understanding of the program and the requirements of the review process.

The program director may also be involved in the selection of panelists for future sessions. This involves identifying potential panelists and evaluating their qualifications and experience.

In conclusion, the program director plays a critical role in the review process. The program director is responsible for ensuring that the review process is conducted fairly and objectively and that the panelists are qualified and experienced.

As we noted, the program director's involvement in the review process is important. The program director must provide oversight and guidance to the panelists. This oversight is necessary to ensure that the panelists are following the guidelines established by the program director and that the review process is conducted fairly and objectively. The program director should also provide feedback to the panelists after the review process. This feedback is important for improving the quality of the review process and for the development of future panelists.

Some program directors may use the same panelists over multiple sessions. This can help to maintain consistency in the review process and to ensure that the panelists have a thorough understanding of the program and the requirements of the review process.

The program director may also be involved in the selection of panelists for future sessions. This involves identifying potential panelists and evaluating their qualifications and experience.

In conclusion, the program director plays a critical role in the review process. The program director is responsible for ensuring that the review process is conducted fairly and objectively and that the panelists are qualified and experienced.
Program Director Activity Prior to Decision Making

29

We present the proposal to a few people before it is finalized. After we have received a few comments, we try to work through them and refine the proposal. Our goal is to ensure that the proposal is as strong as possible before we submit it to the reviewers.

I. Program Director Activity Prior to Decision Making

A. Program Director's Role

The program director is responsible for overseeing the program's activities and ensuring that they meet the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

B. Program Director's Responsibility

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

C. Program Director's Activity

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

D. Program Director's Interaction with Reviewers

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

E. Program Director's Interaction with Applicants

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

F. Program Director's Interaction with Program Directors

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

II. Review Process

A. Review Process Overview

The review process is designed to ensure that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

B. Review Process Details

The review process is designed to ensure that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

C. Review Process Impact

The review process is designed to ensure that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

D. Review Process Challenges

The review process is designed to ensure that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

E. Review Process Solutions

The review process is designed to ensure that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

F. Review Process Outcomes

The review process is designed to ensure that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

III. Conclusion

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

A. Conclusion

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

B. Conclusion

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

C. Conclusion

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

D. Conclusion

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

E. Conclusion

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

F. Conclusion

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

IV. Appendix

A. Appendix

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

B. Appendix

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

C. Appendix

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

D. Appendix

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

E. Appendix

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.

F. Appendix

The program director is responsible for ensuring that the program's activities are aligned with the goals and objectives outlined in the proposal. This includes managing the program's finances, personnel, and resources.
program directors use to obtain knowledgeable reviewers? Among the
variety of sources are: personal knowledge of the field, professional
the NSF offices, journal articles, and proceedings of professional me-
meetings.

I think the selection of nail reviewers is mostly from my personal knowledge of
during the course of the 15 years that I've been teaching and I know people. I
have used in the past.

And:

I have several kinds of lists of reviewers: previously used reviewers. I have
journals and proceedings of professional meetings. I could also look at
the universities.

Locating sources of potential reviewers is apparently somewhat less
difficult than deciding who to ask to review a proposal. A large number
Most program directors tried to get a mix of general and specific
reviews on a proposal. One person put it this way:

Basically, I try to reach guys who are highly qualified in the field. However, I
try to pick one of the people who is not in specialized in that area. He's a familiar
viewpoint.

The need for care in reviewer selection is especially great when work
is being proposed in a somewhat controversial area. Many program
directors talked about this type of case and the ways in which they
handle reviewer selection. One said:

We try to send it to three types of people in these cases. You send it to the
sympathetic ones, knowing their bias. You send it to some known negative
people who are more general, generally competent people who don't fall
on the proposal's strong points.

Another program director spoke of his more general approach to the
problem:

You have to have a good knowledge of the subfield—who are working in
them, what are their conflicts. You have to teach people a calibration on the

Program Director Activity Prior to Decision Making

reviewer. You ask, "What are the conflicts—why this reviewer might or might
not give you a good one."

Sometimes the research areas are so small that there is no way to get
balance among reviewers. One program director described such a
situation:

We had a difficult time for a while because we didn't have people outside that
"school" who were qualified to write proposals. I think a lot of proposals
were dinged in that area and we weren't getting good independent critical
judgments.

The kind of work that the principal investigator is doing also affects
reviewer selection. A physicist program director commented:

We try to make a reasonable balance between experimental people and
theoretical people in the field if there's a significant theoretical component
of the proposal. If it is a very large proposal with the operation in a lab, you try to
select some people who have had experience managing a lab in addition to the
straight physics.

It's not very often that you will ask the average experimentalist to review a
theoretical work, because normally that's not a good idea. However, a very
good experimentalist, with theoretical overtones, will review a theoretical
proposal. But more likely, I will take an experimental proposal and ask a
theoretician to review it because a good theoretician is always looking at
experimental results—that's where he starts from, that's where he leaves off.

Some program directors try to balance industrial and university
people, in fields in which this mix pertains. One director said in this
regard:

We try to get reviews from both industry and from the university. We want to
get advice from practitioners, and depending on the subject matter, we can get
very perceptive damning reviews from some of the industrial people.

Another set of considerations is possible connections between appli-
cants and the reviewers. One program director said:

I check their bibliography to see whether he has any past connection or
collaborative effort with the man or if he was his thesis adviser; so I avoid
people who are colleagues.

The relative eminence of reviewer and proposer is also considered.
A number of program directors spoke of this. One said:

When I get a proposal from a great man, I would use at least two other great
men in reviews. The problem with using young reviewers versus established is
that young reviewers are apt to give innocuous reviews in these cases. A more
experienced reviewer is inclined to say what he thinks one way or another.

A further consideration in selecting reviewers is their "track rec-
ord." One program director said:

There is one man I stopped using because anything I sent him he said was awful
and he didn’t give me any information.

Another expressed a similar view:

There is one type of reviewer that I tend to eliminate—the type that al-
gways gives a negative review and a very low rating. This man will give it a low one,
no matter what it is.

Finally, how reviewers fulfill their obligation to return reviews
affects their selection. One program director said:

Over the years we have built up this list of adequate reviewers on the basis of
the number of times they returned it when you asked them. If you send them
one or two proposals three times a year and you get back one a year, you don’t
send any more.

In sum, considerations relative to selection of reviewers are numer-
ous. Every program director stated that selection of reviewers is
extremely difficult and, also, that it is at the heart of the process.
Directors must have considerable scientific and administrative ex-
pertise to make the kinds of decisions that will lead to the best possible
reviews for proposals. It should also be pointed out that, although the
program directors as a group are aware of the many factors that must
be considered in selecting unbiased reviewers, errors are undoubtedly
made.

Critics claim that the program director can predetermine the outcome
of the peer review process by sending proposals to scientists whose
evaluations of the proposals are predictable. This might be termed the
"old boy hypothesis," which presumes that the proposals of eminent
scientists who are members of the "old boy network" are sent to other
eminent scientists who give the "old boys" favorable evaluations.
Similarly important, the proposals of noneminent scientists, who are not
part of this network, are sent to scientists who will give them lower
evaluations than they deserve. Although we have no evidence one way
or the other that the program directors select reviewers with a certain
outcome in mind, we can, by looking at the outcomes, see whether the
data support the old boy hypothesis. Do eminent reviewers actually
rate the proposals of eminent colleagues more favorably than other
reviewers?5

An immediate problem in testing the old boy hypothesis is the
absence of conceptual clarity in the charge. The charge is that research
money is allocated unfairly, but the attribution of this unfairness to
"old boyism" is somewhat confusing. What is referred to by the label
old boy network? There are at least three possibilities. It could refer to
scientists with a common view of their fields who will favorably
appraise work only by others with similar views. It could refer to social
networks of friendship—made up of scientists who know each other.
“grew up” together, attended the same schools, tend to fraternize, are
of the same sex, and favor each other’s research proposals. It could
TABLE 2 Rank of Department of Reviewers by Rank of Department of Applicants: Algebra

<table>
<thead>
<tr>
<th>Rank of Applicants</th>
<th>Rank of Department of Reviewers, %</th>
<th>Applicants, %</th>
<th>Reviewers, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 15</td>
<td></td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>Other ranked</td>
<td></td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>Unranked and nonacademic</td>
<td></td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>101</td>
<td>101</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank of Applicants</th>
<th>Rank of Department of Reviewers, %</th>
<th>Applicants, %</th>
<th>Reviewers, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 15</td>
<td></td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>Other ranked</td>
<td></td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>Unranked and nonacademic</td>
<td></td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>101</td>
<td>101</td>
</tr>
</tbody>
</table>

1. What is the distribution of applicants and reviewers among different-ranked departments? Are reviewers more or less likely to be drawn from top-ranked departments than are applicants?  
2. Are proposals from applicants currently employed in top-ranked departments more likely to be reviewed by reviewers from top-ranked departments than are proposals from applicants currently employed at less prestigious institutions?  
3. Are applicants from top-ranked departments more likely to receive favorable ratings than are applicants from lower-ranked departments?  
4. Are reviewers from top-ranked departments more or less lenient in their ratings than reviewers from not top-ranked departments?  
5. Are reviewers from top-ranked departments more likely to favor proposals from top-ranked departments than are reviewers from lower-ranked departments?  

In order to respond to these questions, we shall examine in detail the results for 1 of the 10 programs, algebra (see Table 2). The top of the
### TABLE 4: Rank of Department of Reviewers by Rank of Department of Applicants: Biochemistry

<table>
<thead>
<tr>
<th>Rank of Department</th>
<th>Applicants, %</th>
<th>Reviewers, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 15</td>
<td>12</td>
<td>22</td>
</tr>
<tr>
<td>16-32</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>Other ranked</td>
<td>28</td>
<td>14</td>
</tr>
<tr>
<td>Unranked and nonacademic</td>
<td>45</td>
<td>16</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>99</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

### TABLE 5: Rank of Department of Reviewers by Rank of Department of Applicants: Chemical Dynamics

<table>
<thead>
<tr>
<th>Rank of Department</th>
<th>Applicants, %</th>
<th>Reviewers, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 15</td>
<td>13</td>
<td>27</td>
</tr>
<tr>
<td>16-40</td>
<td>25</td>
<td>16</td>
</tr>
<tr>
<td>41-70</td>
<td>30</td>
<td>24</td>
</tr>
<tr>
<td>Unranked and nonacademic</td>
<td>32</td>
<td>35</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>100</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

### TABLE: Rank of Department of Reviewers, Meta Ratings Given

<table>
<thead>
<tr>
<th>Rank of Applicants</th>
<th>Reviewers, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 15</td>
<td>1.79</td>
</tr>
<tr>
<td>16-40</td>
<td>2.05</td>
</tr>
<tr>
<td>41-70</td>
<td>1.82</td>
</tr>
<tr>
<td>Unranked and nonacademic</td>
<td>2.19</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>2.93</strong></td>
</tr>
</tbody>
</table>

---

Table shows the distribution of both applicants and reviewers among different types of departments. In algebra less than one-fifth of the applicants are employed in the 15 top-ranked mathematics departments. A substantial portion (42 percent) of the applicants are currently employed either in departments that are unranked or in nonacademic jobs. The reviewers are far more likely to be in the top-ranked departments than are the applicants. Forty-three percent of the reviewers are currently employed in the top 15 departments.

Do these data by themselves tell us anything about the equity of the review process? The fact that reviewers are drawn heavily from prestigious departments tells us nothing about it. Presumably, there is some concentration of talented applicants in the most prestigious departments.  

1Throughout the analysis we have been forced to use the general disciplinary ratings available in the American Psychological Association (1971). There are no departmental ratings for specialties like algebra. We must therefore assume that high-ranking mathematics departments are in general the most desirable places for algebraists to work.

2Assuming that old baysins is indeed at work, its influence on outcomes is in part dependent upon the extent to which proposals of eminent applicants are disproportionately reviewed by eminent reviewers. The stronger the relationship, the greater the potential influence of old baysins.

---

departments. Many studies have shown that mean faculty prestige and productivity are highly correlated with departmental prestige (Cole and Zuckerman, 1976). Program directors seek reviews, of course, from the best people in the field, and these people tend to be concentrated in the top-ranked departments.

In the second part of Table 2, we report the distribution of ranks of departments of reviewers of proposals from different-ranked departments. In algebra the program director was more likely to assign proposals from applicants in prestigious departments to reviewers in prestigious departments. While 60 percent of the reviewers of proposals from "top-15" applicants were in top-15 departments, only 34 percent of the reviewers of proposals from unranked and nonacademic applicants are located in top-15 departments.
### TABLE 6: Rank of Department of Reviewers by Rank of Department of Applicants: Ecology

<table>
<thead>
<tr>
<th>Rank of Department</th>
<th>Applicants, %</th>
<th>Reviewers, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 10</td>
<td>37</td>
<td>32</td>
</tr>
<tr>
<td>19-50</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>5-0</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Unranked and nonacademic</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>100</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank of Applicants</th>
<th>Rank of Department of Reviewers, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 18</td>
<td>19-50</td>
</tr>
<tr>
<td></td>
<td>Unranked and</td>
</tr>
<tr>
<td></td>
<td>Nonacademic</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>Top 18</td>
<td>33</td>
</tr>
<tr>
<td>19-50</td>
<td>16</td>
</tr>
<tr>
<td>5-0</td>
<td>14</td>
</tr>
<tr>
<td>Unranked and nonacademic</td>
<td>17</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank of Applicants</th>
<th>Rank of Department of Reviewers, Mean Ratings Given</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 18</td>
<td>19-50</td>
</tr>
<tr>
<td></td>
<td>Unranked and</td>
</tr>
<tr>
<td></td>
<td>Nonacademic</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>Top 18</td>
<td>2.58</td>
</tr>
<tr>
<td>19-50</td>
<td>2.22</td>
</tr>
<tr>
<td>5-0</td>
<td>2.18</td>
</tr>
<tr>
<td>Unranked and nonacademic</td>
<td>2.18</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2.16</strong></td>
</tr>
</tbody>
</table>

At first glance, these data may be seen as offering some support for the old boy assumption. Such a conclusion would be erroneous. We do not yet know whether reviewers from top-15 departments are likely to favor applicants from top-15 departments.

The necessary data are presented in the bottom part of Table 2 (low rating = favorable, and high rating = unfavorable). The total column shows that applicants from high-ranked departments are indeed more likely to receive favorable ratings than applicants from lower-ranked departments (comparing the 1.81 with the 2.31). This fits our expectations.

### TABLE 7: Rank of Department of Reviewers by Rank of Department of Applicants: Economics

<table>
<thead>
<tr>
<th>Rank of Department</th>
<th>Applicants, %</th>
<th>Reviewers, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 10</td>
<td>26</td>
<td>55</td>
</tr>
<tr>
<td>Other ranked</td>
<td>29</td>
<td>25</td>
</tr>
<tr>
<td>Unranked and nonacademic</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>100</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank of Applicants</th>
<th>Rank of Department of Reviewers, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 10</td>
<td>19-50</td>
</tr>
<tr>
<td></td>
<td>Unranked and</td>
</tr>
<tr>
<td></td>
<td>Nonacademic</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>Top 10</td>
<td>38</td>
</tr>
<tr>
<td>19-50</td>
<td>24</td>
</tr>
<tr>
<td>5-0</td>
<td>14</td>
</tr>
<tr>
<td>Other ranked</td>
<td>17</td>
</tr>
<tr>
<td>Unranked and nonacademic</td>
<td>100</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank of Applicants</th>
<th>Rank of Department of Reviewers, Mean Ratings Given</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 10</td>
<td>19-50</td>
</tr>
<tr>
<td></td>
<td>Unranked and</td>
</tr>
<tr>
<td></td>
<td>Nonacademic</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>Top 10</td>
<td>1.70</td>
</tr>
<tr>
<td>19-50</td>
<td>1.71</td>
</tr>
<tr>
<td>5-0</td>
<td>2.08</td>
</tr>
<tr>
<td>Other ranked</td>
<td>2.82</td>
</tr>
<tr>
<td>Unranked and nonacademic</td>
<td>2.80</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2.80</strong></td>
</tr>
</tbody>
</table>

top-15 applicants than are other reviewers. The mean review given by reviewers in top-15 departments to proposals submitted by applicants in top-15 departments is 1.98. This is a less favorable score than the mean review given by reviewers in lower-ranked departments. The information from Table 2 is summarized in the first row of Table 12. The first statistical test performed is a comparison of the mean rating of applicants from the top group of departments with the general mean rating. As we see in algebra, this is statistically significant at the 0.005 level. The figures in this part of the table simply tell us the following.

The next section of Table 12 indicates whether the reviewers from top-ranked departments are more likely to be lenient or tough than are reviewers from other departments. Since the mean rating given by top-ranked reviewers in departments is lower than the mean rating given by all reviewers and the difference is statistically significant, we can
conclude that, in algebra, top-ranked reviewers are more lenient, in general, than reviewers from other departments. This is the case in only 4 of the 10 programs: algebra, chemical dynamics, economics, and solid-state physics. In the other six programs, top-ranked reviewers were in general less lenient than were other reviewers.

The last part of the table shows whether there was any significant interaction effect. That is, are top-ranked reviewers more likely to give high scores to applicants from top-ranked departments than would be expected on the basis of the general tendency of top-ranked reviewers to give low scores and the general tendency of top-ranked applicants to get low scores? Given these two distributions, the expected mean rating of top-15 applicants in algebra would be 1.68, and the observed mean rating was 1.98. Thus, there is no evidence that old bosun is at work in this program. In fact, the data show that if anything, top-ranked people are tougher on their colleagues at top-ranked institutions than would be expected.

The last column of Table 12 shows that only in biochemistry was there any statistically significant interaction. That is, only in biochemistry are reviewers from top-ranked departments more likely to give favorable ratings to applicants from top-ranked departments than would be expected by chance. This could indicate some degree of bias or that in this field it is possible that reviewers in top-ranked departments are more discriminating and are more able to assess high-quality proposals. In seven of the programs, the relationship had effects opposite to those expected; that is, top-ranked reviewers gave lower scores to proposals from top-ranked applicants than would be expected by chance. In the two other programs, anthropology and meteorology, the differences were not statistically significant. On the basis of these data, there is very little evidence that reviewers were biased in evaluating the proposals of their colleagues.

We also considered the effect of the geographic location of reviewers on how they evaluated proposals from applicants in different geographic locations. The results are presented in Table 13. The first column presents the mean rating given where the geographic location...
<table>
<thead>
<tr>
<th>Rank of Department</th>
<th>Applicants, %</th>
<th>Reviewers, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 17</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>Other ranked</td>
<td>33</td>
<td>39</td>
</tr>
<tr>
<td>Unranked and nonacademic</td>
<td>42</td>
<td>46</td>
</tr>
<tr>
<td>TOTAL</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank of Department</th>
<th>Applicants, %</th>
<th>Reviewers, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 17</td>
<td>19</td>
<td>13</td>
</tr>
<tr>
<td>Other ranked</td>
<td>37</td>
<td>35</td>
</tr>
<tr>
<td>Unranked and nonacademic</td>
<td>43</td>
<td>40</td>
</tr>
<tr>
<td>TOTAL</td>
<td>99</td>
<td>99</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank of Department</th>
<th>Applicants, %</th>
<th>Reviewers, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 17</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Other ranked</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>Unranked and nonacademic</td>
<td>43</td>
<td>100</td>
</tr>
<tr>
<td>TOTAL</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank of Department</th>
<th>Applicants, %</th>
<th>Reviewers, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 17</td>
<td>2.79</td>
<td>2.98</td>
</tr>
<tr>
<td>Other ranked</td>
<td>2.87</td>
<td>2.57</td>
</tr>
<tr>
<td>Unranked and nonacademic</td>
<td>3.34</td>
<td>3.00</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2.50</td>
<td>2.82</td>
</tr>
</tbody>
</table>

*Rank based upon judgment of program director.*

The second column shows the mean rating given when the geographic location of the applicant and reviewer are not the same. If the number in column 1 is higher than the number in column 2, there is no evidence that reviewers are, in general, more likely to favor people from the same part of the country. For each of the 10 fields, the relationship produces effects opposite to those expected; that is, reviewers are more harsh on proposals from people in their own areas than they are on proposals from people in other areas. In the three other areas, although the relationship is as expected, it is statistically nonsignificant.

We then tested four separate hypotheses related to whether reviewers were likely to favor applicants from their own areas. For example, the mean ratings given to applicants from the northeast by reviewers from the northeast (column 4) can be compared with the mean ratings given to applicants from the northeast by reviewers from other sections of the country (column 5). Once again, the number in column 4 would have to be lower than the number in column 5 to demonstrate regional bias in reviewing. We do find a statistically significant relationship in 2 of the 10 programs, fluid mechanics and meteorology. In these two programs, reviewers from the northeast are more lenient on proposals from applicants from the northeast than are reviewers from other sections of the country.

Proposals from southerners show no evidence of any regional bias in any of the 10 programs. They are given the same evaluations by southerners as by reviewers from other sections of the country. In all, we made 30 such geographic comparisons. In only six cases did we find statistically significant differences. There is very little evidence that reviewers in certain geographic locations rate the proposals of applicants in those locations more favorably than do reviewers from other sections of the country.

For one field, biochemistry, we collected data on the citations of the reviewers. We found no correlation between numbers of citations of the applicant and reviewer are the same. The second column shows the mean rating given when the geographic location of the applicant and reviewer are not the same. If the number in column 1 is higher than the number in column 2, there is no evidence that reviewers are, in general, more likely to favor people from the same part of the country. For each of the 10 fields, the relationship produces effects opposite to those expected; that is, reviewers are more harsh on proposals from people in their own areas than they are on proposals from people in other areas. In the three other areas, although the relationship is as expected, it is statistically nonsignificant.

We then tested four separate hypotheses related to whether reviewers were likely to favor applicants from their own areas. For example, the mean ratings given to applicants from the northeast by reviewers from the northeast (column 4) can be compared with the mean ratings given to applicants from the northeast by reviewers from other sections of the country (column 5). Once again, the number in column 4 would have to be lower than the number in column 5 to demonstrate regional bias in reviewing. We do find a statistically significant relationship in 2 of the 10 programs, fluid mechanics and meteorology. In these two programs, reviewers from the northeast are more lenient on proposals from applicants from the northeast than are reviewers from other sections of the country.

Proposals from southerners show no evidence of any regional bias in any of the 10 programs. They are given the same evaluations by southerners as by reviewers from other sections of the country. In all, we made 30 such geographic comparisons. In only six cases did we find statistically significant differences. There is very little evidence that reviewers in certain geographic locations rate the proposals of applicants in those locations more favorably than do reviewers from other sections of the country.

For one field, biochemistry, we collected data on the citations of the reviewers. We found no correlation between numbers of citations of the applicant and reviewer are the same. The second column shows the mean rating given when the geographic location of the applicant and reviewer are not the same. If the number in column 1 is higher than the number in column 2, there is no evidence that reviewers are, in general, more likely to favor people from the same part of the country. For each of the 10 fields, the relationship produces effects opposite to those expected; that is, reviewers are more harsh on proposals from people in their own areas than they are on proposals from people in other areas. In the three other areas, although the relationship is as expected, it is statistically nonsignificant.

We then tested four separate hypotheses related to whether reviewers were likely to favor applicants from their own areas. For example, the mean ratings given to applicants from the northeast by reviewers from the northeast (column 4) can be compared with the mean ratings given to applicants from the northeast by reviewers from other sections of the country (column 5). Once again, the number in column 4 would have to be lower than the number in column 5 to demonstrate regional bias in reviewing. We do find a statistically significant relationship in 2 of the 10 programs, fluid mechanics and meteorology. In these two programs, reviewers from the northeast are more lenient on proposals from applicants from the northeast than are reviewers from other sections of the country.

Proposals from southerners show no evidence of any regional bias in any of the 10 programs. They are given the same evaluations by southerners as by reviewers from other sections of the country. In all, we made 30 such geographic comparisons. In only six cases did we find statistically significant differences. There is very little evidence that reviewers in certain geographic locations rate the proposals of applicants in those locations more favorably than do reviewers from other sections of the country.

For one field, biochemistry, we collected data on the citations of the reviewers. We found no correlation between numbers of citations of the applicant and reviewer are the same. The second column shows the mean rating given when the geographic location of the applicant and reviewer are not the same. If the number in column 1 is higher than the number in column 2, there is no evidence that reviewers are, in general, more likely to favor people from the same part of the country. For each of the 10 fields, the relationship produces effects opposite to those expected; that is, reviewers are more harsh on proposals from people in their own areas than they are on proposals from people in other areas. In the three other areas, although the relationship is as expected, it is statistically nonsignificant.

We then tested four separate hypotheses related to whether reviewers were likely to favor applicants from their own areas. For example, the mean ratings given to applicants from the northeast by reviewers from the northeast (column 4) can be compared with the mean ratings given to applicants from the northeast by reviewers from other sections of the country (column 5). Once again, the number in column 4 would have to be lower than the number in column 5 to demonstrate regional bias in reviewing. We do find a statistically significant relationship in 2 of the 10 programs, fluid mechanics and meteorology. In these two programs, reviewers from the northeast are more lenient on proposals from applicants from the northeast than are reviewers from other sections of the country.

Proposals from southerners show no evidence of any regional bias in any of the 10 programs. They are given the same evaluations by southerners as by reviewers from other sections of the country. In all, we made 30 such geographic comparisons. In only six cases did we find statistically significant differences. There is very little evidence that reviewers in certain geographic locations rate the proposals of applicants in those locations more favorably than do reviewers from other sections of the country.

For one field, biochemistry, we collected data on the citations of the reviewers. We found no correlation between numbers of citations of the applicant and reviewer are the same. The second column shows the mean rating given when the geographic location of the applicant and reviewer are not the same. If the number in column 1 is higher than the number in column 2, there is no evidence that reviewers are, in general, more likely to favor people from the same part of the country. For each of the 10 fields, the relationship produces effects opposite to those expected; that is, reviewers are more harsh on proposals from people in their own areas than they are on proposals from people in other areas. In the three other areas, although the relationship is as expected, it is statistically nonsignificant.

We then tested four separate hypotheses related to whether reviewers were likely to favor applicants from their own areas. For example, the mean ratings given to applicants from the northeast by reviewers from the northeast (column 4) can be compared with the mean ratings given to applicants from the northeast by reviewers from other sections of the country (column 5). Once again, the number in column 4 would have to be lower than the number in column 5 to demonstrate regional bias in reviewing. We do find a statistically significant relationship in 2 of the 10 programs, fluid mechanics and meteorology. In these two programs, reviewers from the northeast are more lenient on proposals from applicants from the northeast than are reviewers from other sections of the country.

Proposals from southerners show no evidence of any regional bias in any of the 10 programs. They are given the same evaluations by southerners as by reviewers from other sections of the country. In all, we made 30 such geographic comparisons. In only six cases did we find statistically significant differences. There is very little evidence that reviewers in certain geographic locations rate the proposals of applicants in those locations more favorably than do reviewers from other sections of the country.

For one field, biochemistry, we collected data on the citations of the reviewers. We found no correlation between numbers of citations of
<table>
<thead>
<tr>
<th>NS</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>600</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>700</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>800</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>900</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
reviewers and numbers of citations of applicants. The program director was not more likely to assign the proposals of eminent biochemists for review by other eminent biochemists, at least as eminence is measured by citations. We also found no evidence that reviewers with many citations were more likely to be lenient than were reviewers with fewer citations. As expected, however, applicants with relatively large numbers of citations to their recent work, in general, received significantly more favorable reviews than applicants with relatively few citations. Finally, and most importantly, the interaction effect is statistically insignificant. Thus, there is no evidence that highly cited reviewers are excessively favorable to the proposals of highly cited applicants.

In conclusion, we find little evidence that the characteristics of reviewers interact with the characteristics of applicants so as to influence substantively the outcome of decisions. Different types of reviewers seem to evaluate proposals of different types of applicants in much the same way. There is very little evidence for reviewer bias or for support of an old boy hypothesis. We must collect additional data to test the other forms of that hypothesis.

SECTION 3

Influence of Characteristics of Applicants on Reviewer Ratings

We have demonstrated that the ratings received by applicants are not significantly influenced by the characteristics of the scientists doing the rating—the peer reviewers. We now examine another question. To what extent are the ratings given by peer reviewers correlated with the characteristics of principal investigators, or applicants? Four essential criteria are supposed to be applied in the evaluation of applications:

1. The quality of science described in the proposal.
2. The competence of the principal investigator to conduct the research as demonstrated by past scientific performance.
3. Available facilities.
4. Geographic and institutional distribution, all other things being equal.

We can use the data on the reviews received by the 1,200 applicants we have studied to see the extent to which favorable ratings are more likely to be received by scientists in the most prestigious institutions and by scientists who have been funded by the NSF in the past.

The quantitative analysis reported in this section has two different purposes. The first is to discover the extent to which the characteristics of NSF applicants, including measures of their past scientific research performance, can be used to predict the ratings their proposals receive from peer reviewers. The second is to discover the extent to which different types of applicants with distinctly different characteristics are
TABLE 14: Frequency Distribution of Ratings for Each Program

In order to demonstrate the difference between the two questions, consider a concrete example. The first question, involving comparison of ratings for different types of scientists, asks: Are the applicants whose work has more published citations or very good ratings in the bottom 10 percent? The second question, involving comparison of ratings for different types of applications, asks: Are the proposals with more published citations or very good ratings in the bottom 10 percent? The second question is easier to answer because the criteria for comparing the two types of applications are set forth in the first question. All applications should be evaluated using the same mathematical criteria. The second question requires the analyst to choose an application type and then evaluate it using the same standards as the first question. In answering the second question, we depend on tables such as Table 14, which shows the frequency distribution of ratings for each program.
TABLE 15: Mean and standard deviations for all 104 combinations.

<table>
<thead>
<tr>
<th>Track Record</th>
<th>Peer Review Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>Intelligence</td>
</tr>
<tr>
<td>Young Scientists</td>
<td>Published papers</td>
</tr>
<tr>
<td>Published papers</td>
<td>Other factors</td>
</tr>
</tbody>
</table>

The table shows the mean and standard deviations for all 104 combinations of factors. The factors include track record, peer review ratings, performance, intelligence, young scientists, published papers, other factors, and their combinations. The table provides a comprehensive view of the statistically significant factors influencing the success of scientists.
Influence of Characteristics of Applicants on Ratings

We have used simple ordinary least-squared regression analysis here. The cell entries of Table 16 present the squared zero-order correlation coefficient, which is simply the proportion of variance on the dependent variable, ratings, explained by the independent variable, citations to recent work. The higher the numbers the more variance on the ratings can be explained by citations to recent work. In all fields except anthropology, citations to recent work explain some variance on the ratings received. The most interesting fact about these data, however, is that citations to past work explain so little variance in the ratings. Even in biochemistry and chemical dynamics, in which citations explain the most variance in ratings received, they explain less than a fifth of the variance; in most fields, they explain considerably less. This means that scientists who have demonstrated their competence by publishing frequently cited papers are more likely to receive favorable ratings but that this effect is weak. In fact, the great bulk of the variance in the ratings cannot be explained by citations to recent work.

We examined not only the 10 fields separately, but also all 10 programs combined. Since the mean and standard deviations on the relevant variables differ significantly from program to program, it is necessary first to standardize separately all the data within each field before combining data on applicants from different programs. For

**Table 16: Proportion of Variance Explained ($R^2$) on Rating by Citations to Recent Work: 10 Programs**

<table>
<thead>
<tr>
<th>Subject</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebra</td>
<td>0.06</td>
</tr>
<tr>
<td>Anthropology</td>
<td>0.00</td>
</tr>
<tr>
<td>Biochemistry</td>
<td>0.16</td>
</tr>
<tr>
<td>Chemical Dynamics</td>
<td>0.14</td>
</tr>
<tr>
<td>Ecology</td>
<td>0.01</td>
</tr>
<tr>
<td>Economics</td>
<td>0.08</td>
</tr>
<tr>
<td>Fluid Mechanics</td>
<td>0.03</td>
</tr>
<tr>
<td>Geophysics</td>
<td>0.07</td>
</tr>
<tr>
<td>Meteorology</td>
<td>0.08</td>
</tr>
<tr>
<td>Solid State Physics</td>
<td>0.08</td>
</tr>
</tbody>
</table>

*Log of citations made in 1975 to work published between 1965 and 1974.*

For a description of this index see Appendix B. We have also used log transformations on the number of papers published, since this variable is also highly skewed.

*For a complete bibliography of studies using citations see any one of the annual guides published by the Institute for Scientific Information to accompany the Science Citation Index.*
example, we will express the number of citations received by a biochemist not in terms of an absolute number but rather in terms of the number of standard deviations above or below the mean for biochemists. Thus, a biochemist who is one standard deviation above the mean for biochemistry in citations would be treated as equivalent to an anthropologist who is one standard deviation above the mean for anthropology, despite the fact that the biochemist would have many more citations than the anthropologist. We are converting absolute scores on the variables into scores relative to other individuals in the same program. These relative or standardized scores are comparable across programs.

After standardizing the data separately within each field we were able to treat all pairs of reviewers and applicants as one sample. When we use the standardized data for all 10 fields combined, we find that citations made in 1974 to work published between 1965 and 1974 explain 6 percent of the variance in ratings.

In order to explicate still further the meaning of the results we have obtained from the regression analysis we present several scattergrams displaying the relationship between the selected variables and ratings. We begin by looking at the relationship between citations to recent work and ratings for the entire sample combined. (See Figure 3.)

The cloud of points in Figure 3 indicates that there is not a strong relationship between the two variables being plotted. If citations were a good predictor of the ratings received by applicants on their proposals, we should expect that those applicants who had high citation scores, that is, were located at the far right of the scattergram, would be heavily clustered in the lower part of the scattergram, indicating that they had received "low" numerical but "high" adjectival ratings. (The reader must always invert these scores in his mind, since the SRF codes are "excellent" as a 1 and a "poor" as a 5.) Between these are 2, "very good," 3, "good," and 4, "fair." We would also expect to find those scientists who had received few or no citations, those appearing in the far left of the scattergram, clustered in the top half of the scattergram, indicating that they had received relatively low ratings on the proposals. This is clearly not the case. A substantial portion of the ratings of scientists with relatively large numbers of citations are relatively high (read low). Scientists with relatively few citations to their recent work

*An asterisk in Figure 5 indicates that one reviewer and applicant pair was located in this particular point in the scatterplot. Numbers 2 and 4 indicate the number of different pairs of citations located at those locations. The computer program used to generate these scatterplots did not have the capability of indicating a large number of citations at any particular location. Therefore, a number 9 indicates that 9 or more pairs were at this location in the scattergram.
received relatively low (read high) ratings on their proposals. In other words, there is substantial overlap in the ratings received by highly cited scientists and those with few citations. Thus it is impossible to predict accurately the rating a scientist's proposal will get from knowledge of the number of citations to the recent work of that scientist.

We illustrate further the lack of a relationship between citations to recent work and peer review ratings by considering the results obtained for 2 of the 10 programs: biochemistry, in which the association between these 2 variables was highest among the 10 programs, and ecology, which had the second lowest association. Figure 4 presents the scattergram for the relationship between citations to recent work and ratings received in biochemistry. Again, a large number of scientists whose recent work has received a substantial number of citations obtained relatively poor peer review ratings. Conversely, many scientists who have received few citations to their recent work obtained very good or excellent peer review ratings, represented by "low" scores on the top rating scale. In short, this scattergram suggests that the association between these two variables is relatively weak. This is even more apparent when we examine Figure 5, in which we present the same relationship for applicants to the ecology program. Here we see almost no relationship between these two variables.

Thus far we have used two analytic techniques to explore the possibility that a scientist's past track record is associated with peer review ratings. At least for this one indicator of track record, we have concluded that there is no substantial relationship between ratings and citations to recent work. In fact, using simple regression models we find a very pronounced lack of fit between the data and the model. Examination of the scattergrams suggests why the regression model does not provide a good description of the relationship between citations to recent work and peer review ratings. It is unlikely that any simple function could describe the data presented here.

Now we compare the results obtained from regression and scattergram analyses with those obtained from tabular analysis of the same data. In Table 17 we show the relationship between the number of citations received in 1974 to work published between 1965 and 1974 and the rating received on the proposal. For purposes of tabular analysis we dichotomized ratings into excellent or very good (the two highest rating categories) and all others. Thus, within each program we show for each citation category the proportion of applicants who received excellent or very good ratings. For example, in 134 cases in algebra the applicant had no citations. In 65 percent of these cases the applicant received a rating of excellent or very good. In tabular
<table>
<thead>
<tr>
<th>Coordinates on edge of pavement</th>
<th>25 m</th>
<th>30 m</th>
<th>35 m</th>
<th>40 m</th>
<th>45 m</th>
<th>50 m</th>
<th>55 m</th>
<th>60 m</th>
<th>65 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>North-West</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>North-East</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>South-West</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>South-East</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
<td>0.9</td>
</tr>
</tbody>
</table>

**TABLE 1**

Application of Concrete Overlay on Road Network in [year] to Reduce Pavement

---

*Please note that the table above is a representation of the data provided in the image. The actual data and its context should be verified from the source.*
Influence of Characteristics of Applicants on Rating

Solid-state physics is a field in which the relationship between citations and ratings is modest; a total of 60 percent of the ratings were either excellent or very good, with the number of citations.

Some applicants were categorized into the independent variable, while others were categorized into the dependent variable.

The data on solid-state physics in Table 17 are given in the form of a table showing the number of citations and the number of ratings at each level.

The proportion of those who received high ratings is approximately 70 percent, while the proportion of those who received low ratings is approximately 30 percent.

The distribution of citations and ratings is as follows:

- Excellent: 30 percent
- Very good: 70 percent
- Good: 0 percent
- Fair: 0 percent
- Poor: 0 percent

The distribution of citations is as follows:

- Excellent: 0 percent
- Very good: 0 percent
- Good: 0 percent
- Fair: 0 percent
- Poor: 0 percent

The distribution of ratings is as follows:

- Excellent: 0 percent
- Very good: 0 percent
- Good: 0 percent
- Fair: 0 percent
- Poor: 0 percent

The distribution of citations and ratings is as follows:

- Excellent: 0 percent
- Very good: 0 percent
- Good: 0 percent
- Fair: 0 percent
- Poor: 0 percent

The relationship between citations and ratings is not significant.

The proportion of those who received high citations and ratings is approximately 70 percent, while the proportion of those who received low citations and ratings is approximately 30 percent.

The distribution of citations is as follows:

- Excellent: 0 percent
- Very good: 0 percent
- Good: 0 percent
- Fair: 0 percent
- Poor: 0 percent

The distribution of ratings is as follows:

- Excellent: 0 percent
- Very good: 0 percent
- Good: 0 percent
- Fair: 0 percent
- Poor: 0 percent

The distribution of citations and ratings is as follows:

- Excellent: 0 percent
- Very good: 0 percent
- Good: 0 percent
- Fair: 0 percent
- Poor: 0 percent

The distribution of citations and ratings is as follows:

- Excellent: 0 percent
- Very good: 0 percent
- Good: 0 percent
- Fair: 0 percent
- Poor: 0 percent

The distribution of citations and ratings is as follows:

- Excellent: 0 percent
- Very good: 0 percent
- Good: 0 percent
- Fair: 0 percent
- Poor: 0 percent

The relationship between citations and ratings is not significant.

The proportion of those who received high citations and ratings is approximately 70 percent, while the proportion of those who received low citations and ratings is approximately 30 percent.

The distribution of citations is as follows:

- Excellent: 0 percent
- Very good: 0 percent
- Good: 0 percent
- Fair: 0 percent
- Poor: 0 percent

The distribution of ratings is as follows:

- Excellent: 0 percent
- Very good: 0 percent
- Good: 0 percent
- Fair: 0 percent
- Poor: 0 percent

The distribution of citations and ratings is as follows:

- Excellent: 0 percent
- Very good: 0 percent
- Good: 0 percent
- Fair: 0 percent
- Poor: 0 percent

The distribution of citations and ratings is as follows:

- Excellent: 0 percent
- Very good: 0 percent
- Good: 0 percent
- Fair: 0 percent
- Poor: 0 percent

The relationship between citations and ratings is not significant.
Influence of Characteristics of Applicants on Ratings

TABLE 18  Proportion of Variance Explained (R²) on Rating by Citations to Old Works: 10 Programs

<table>
<thead>
<tr>
<th>Field</th>
<th>Explained R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algèbre</td>
<td>0.03</td>
</tr>
<tr>
<td>Anthropology</td>
<td>0.02</td>
</tr>
<tr>
<td>Biochemistry</td>
<td>0.06</td>
</tr>
<tr>
<td>Chemie-Physik</td>
<td>0.05</td>
</tr>
<tr>
<td>Ecology</td>
<td>0.01</td>
</tr>
<tr>
<td>Economics</td>
<td>0.02</td>
</tr>
<tr>
<td>Fluid Mechanics</td>
<td>0.07</td>
</tr>
<tr>
<td>Geophysics</td>
<td>0.02</td>
</tr>
<tr>
<td>Meteorology</td>
<td>0.01</td>
</tr>
<tr>
<td>Solid-State Physics</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Log of citations made in 1974 to work published prior to 1960.

Published 10 or more years ago. When we use the standardized data for all 10 programs combined, we find that the citations to older work explain only 2 percent of the variance on rating. The tabular data on this variable are presented in Table 19.

Table 20 presents the proportion of variance explained on ratings by the numbers of papers published in the last 10 years, 14 of the 10 programs (algèbre, anthropology, ecology, and economics) and no variance is explained. The number of papers published explains only 1 percent of the variance in ratings in fluid mechanics, 7 percent for biochemistry, 9 percent for chemical dynamics, and 12 percent for meteorology. When we use the standardized data for all 10 programs combined, we find that the number of papers published in 1965-1974 explains 2 percent of the variance in ratings.

It is worth noting that there are generally high correlations between the number of papers a scientist has published and the number of times that he or she has been cited. In fact, in biochemistry, the field that on the average shows the highest correlation between the three productivity variables and the ratings, we found that all three variables together explained only 17 percent of the variance, only 1 percent more than explained by citations to recent work.

Table 21 presents the relationship between the total number of papers published between 1965 and 1974 and ratings received by applicants in each of the 10 fields. In 7 of the 10 programs there is less than a 20 percent point difference between scientists who have published more than 10 papers and those who have published less than 10 papers.
The characteristics of the papers published between 1965 and 1974, as rated by panelists, showed a significant variation. The table below provides a summary of the ratings:

<table>
<thead>
<tr>
<th>Authorship</th>
<th>Language</th>
<th>Results</th>
<th>Discussion</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.97</td>
<td>0.07</td>
<td>0.99</td>
<td>0.74</td>
</tr>
</tbody>
</table>

It is clear that authors from different countries have varying degrees of influence in the academic community. The table above highlights the disparities in publication results and the level of discussion and reference citation.

In conclusion, the results indicate that international collaboration and the quality of research material significantly impact academic publication success.
discussion at end of this section.) The reviewing process could contain a large arbitrary element. If this is the case, we will find a low correlation between ratings given by the year reviewers and ratings given by independently chosen sets of reviewers. Phase 2 of this research project, which is currently under way, will investigate this possibility.

We are concerned with one other variable as an indicator of the past track record of principal investigatorsthe number of years out of the last 5 in which they have received NSF funds. Some applicants had received NSF funds in all or several of the years, whereas others had received NSF funds in none of those 5 years. Do applicants who currently are or recently have been NSF grant recipients have a greater likelihood of getting favorable ratings from reviewers? The data in Table 22 indicate that whether or not applicants are recent past recipients of NSF funds has very little influence on ratings of their current applications. In all 10 programs the proportion of variance explained by funding history is low. In one program it is 0. In two others it is 1 percent of the variance, and in two others it is only 2 percent of the variance. The greatest proportion of variance explained is in Economics, but even here only 6 percent of the variance is explained by funding histories of applicants. When we use the standardized data for all 10 programs combined, we find that recent NSF funding history explains 5 percent of the variance in rating. Again, we conclude that recent NSF funding history has relatively little influence on the ratings received.

In Figure 6 we present a scattergram displaying the relationship between the funding history of applicants and ratings received on their current proposals. For this scattergram we have used the combined sample of standardized data for all 10 programs. As we would expect

TABLE 22 Proportion of Variance Explained (R²) on Ratings by Year Funded: 1970-1974

<table>
<thead>
<tr>
<th>Subject</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebra</td>
<td>0.05</td>
</tr>
<tr>
<td>Anthropology</td>
<td>0.09</td>
</tr>
<tr>
<td>Biochemistry</td>
<td>0.05</td>
</tr>
<tr>
<td>Chemical Dynamics</td>
<td>0.02</td>
</tr>
<tr>
<td>Ecology</td>
<td>0.08</td>
</tr>
<tr>
<td>Economics</td>
<td>0.05</td>
</tr>
<tr>
<td>Fluid Mechanics</td>
<td>0.01</td>
</tr>
<tr>
<td>Geography</td>
<td>0.03</td>
</tr>
<tr>
<td>Meteorology</td>
<td>0.06</td>
</tr>
<tr>
<td>Solid State Physics</td>
<td>0.02</td>
</tr>
</tbody>
</table>
from the previous associations just reported, the scattergram shows that there is no significant association between the two variables. Many applicants who have received no war funding in the last 5 years received favorable ratings on their proposals, and many applicants who have been funded during the entire period received unfavorable ratings on their proposals. In short, knowledge of whether or not an applicant has been funded by the war in the recent past is of little or no use in predicting the rating of a current proposal. It is clear from the cloud of points presented in Figure 6 why the regression results of Table 22 show no significant association between granting history and ratings received.

Table 23 shows the relationship between granting history and ratings received, using tabular analysis. For the tabular analysis we have dichotomized the applicants into those who have and those who have not received war funds in the last 5 years. In one program, anthropology, applicants who recently received war funds actually had a slightly higher probability of getting excellent or very good ratings on their proposals than did applicants who had not received war funds in the last 5 years. In all the other nine programs the differences between the two groups of applicants in the proportion receiving excellent or very good ratings were only slight to moderate but are definitely worth noting. The field showing the strongest relationship was economics. In this field, 73 percent of past war grantees and 42 percent of those who had not received war funds received excellent or very good ratings on their proposals.

### TABLE 23 Applicants Receiving Excellent or Very Good Ratings by Past Funding History: 10 Programs

<table>
<thead>
<tr>
<th>Program</th>
<th>Received NSF Funds in Last 5 Years, %</th>
<th>Did Not Receive NSF Funds in Last 5 Years, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebra</td>
<td>71 (152)</td>
<td>66 (131)</td>
</tr>
<tr>
<td>Anthropology</td>
<td>48 (48)</td>
<td>53 (169)</td>
</tr>
<tr>
<td>Biochemistry</td>
<td>58 (298)</td>
<td>37 (172)</td>
</tr>
<tr>
<td>Chemical Dynamics</td>
<td>74 (166)</td>
<td>56 (187)</td>
</tr>
<tr>
<td>Ecology</td>
<td>67 (181)</td>
<td>55 (204)</td>
</tr>
<tr>
<td>Economics</td>
<td>33 (95)</td>
<td>42 (214)</td>
</tr>
<tr>
<td>Fluid Mechanics</td>
<td>51 (174)</td>
<td>44 (151)</td>
</tr>
<tr>
<td>Geophysics</td>
<td>44 (297)</td>
<td>53 (181)</td>
</tr>
<tr>
<td>Meteorology</td>
<td>45 (251)</td>
<td>34 (222)</td>
</tr>
<tr>
<td>Solid State Physics</td>
<td>7 (225)</td>
<td>57 (267)</td>
</tr>
</tbody>
</table>

Numbers of applicants in parentheses.
LOCATION AT PRESTIGIOUS DEPARTMENTS AND PEER REVIEW RATINGS

Our data also tell us whether peer reviewers are more likely to give favorable ratings to scientists in the most prestigious academic departments. We might expect to find some correlation, since presumably some departments are more highly ranked than others because they have more superior scientists in them. These scientists should get higher ratings both because of their capabilities as scientists and because it is presumed that their research proposals are better. As the data in Table 24 show, however, there is not a strong correlation between the rank of an applicant's current department and the rating he receives from peer reviewers. In all the programs, with the exception of anthropology, there is a correlation between the rank of applicants' departments and the ratings given their proposals; but again these correlations are surprisingly low. The greatest proportion of variance explained by department rankings is in economics, but even here only 13 percent of the variance is explained by department ranking. These data lead to the conclusion that reviewers are not being significantly influenced by the affiliations of applicants. They are only slightly more apt to give higher ratings to applicants from prestigious institutions than to those from less prestigious institutions. When we use the standardized data for all 10 programs combined, we find that rank of current department explains 5 percent of the variance in ratings.

Figure 7 shows what the relationship would be between the rank of an applicant's department and the applicant's rating if there were a

TABLE 24: Proportion of Variance Explained ($R^2$) on Rating by Rank of Present Department

<table>
<thead>
<tr>
<th>Major</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebra</td>
<td>0.07</td>
</tr>
<tr>
<td>Anthropology</td>
<td>0.00</td>
</tr>
<tr>
<td>Biochemistry</td>
<td>0.07</td>
</tr>
<tr>
<td>Chemical Engineering</td>
<td>0.02</td>
</tr>
<tr>
<td>Ecology</td>
<td>0.02</td>
</tr>
<tr>
<td>Economics</td>
<td>0.13</td>
</tr>
<tr>
<td>Fluid Mechanics</td>
<td>0.10</td>
</tr>
<tr>
<td>Geophysics</td>
<td>0.03</td>
</tr>
<tr>
<td>Meteorology</td>
<td>0.05</td>
</tr>
<tr>
<td>Solid State Physics</td>
<td>0.08</td>
</tr>
</tbody>
</table>

$^*$Rank of department score based upon survey of NAS members (see Appendix B).
perfect or maximum association between these two variables. For this illustration we use the standardized data from the combined sample. We have roughly drawn in the least-squares regression line. The slope is 0.904. (Since the data are standardized, the correlation coefficient is 1 because rank of department is not a continuous variable.) If all the highest-rank departments were assigned to the applicants from the highest-rank departments, only 82 percent of the variance in ratings would be explained.

The actual distribution of the data is presented in Figure 8. The slope, or the correlation coefficient, for this regression line is 0.215. Thus, the rank of an applicant's department explains only about 0.045 percent of the variance on ratings; the regression equation is an inadquate predictor of an applicant's rating. Examining the extent to which the points in the scattergram are spread out both above and below the regression line and the tremendous amount of overlap in the ratings scores for people in departments of different rank emphasizes the importance of other factors. Table 23 shows the relationship between the rank of an applicant's current department and the ratings received on the proposal. Applicants from highest-ranked departments are indeed more likely to receive higher ratings than those from unranked or nonacademic departments. However, in several of the programs, such as chemical dynamics, the relationship is very weak. In ecology and geophysics, the relationship is nonexistent. For example, in several of the programs the relationship is negative. As a result of these negative relationships, the distribution of ratings for people in the next group received high ratings, but 78 percent of those in the fourth group received high ratings—the same proportion as that received by scientists in the highest-ranked departments. Other factors show a similar lack of linearity. For example, in solid state physics, scientists located in the lowest-ranked departments received just about the same proportion of high ratings as those in the most prestigious departments.

These findings on the relationship between rank of department and ratings seem to contradict common sense. On closer examination, ratings were not as consistent as in the previous studies. However, they corroborate the findings of prior empirical studies in the sociology of science. Although it is true that, on the average, highly prestigious departments have more productive and talented scientists, a non-negligible proportion of talented scientists are not in the most prestigious departments. Several independent studies have found that the correlation between citations to a scientist's work and the prestige...
Influence of Characteristics of Applicants on Ratings | 75

rank of his department is 0.30 or less. (For a review of the literature on this topic see Cole and Zucker, 1976.) This means that quite a few scientists who have produced high-quality work are not in highly ranked departments.

When we relate the low correlation between the quality of an individual scientist's research output and the rank of his department to the concept of self-selection we can understand better the low correlation between the rank of an applicant's department and peer review ratings. If every scientist in every department applied for a grant, there would probably be a considerably higher correlation between rank of department and rating. But we know that all scientists do not apply. Applying scientists from low-ranked departments are probably the most active researchers. Whereas six mathematicians from MIT may apply for NSF funds in a given year, perhaps only one mathematician at a lower-ranked department will apply. But this one man will possibly have a national reputation comparable to some of those of his colleagues at higher-ranked departments. The relatively wide dispersion of scientific talent and the process of self-selection may well provide the explanation of the data in Table 25.

To illustrate how tabular analysis allows us to compare people at the extremes of a distribution, we have computed an index in which applicants are given scores based upon the quintiles of their citations and the quintile ranks of their current departments. A scientist in the highest-ranked department with the highest number of citations would receive a score of 10. A scientist in the lowest-ranked department with the lowest number of citations would receive a score of 0. Table 26 shows the proportion of applicants in each index category who received excellent or very good ratings. Thirty-four percent of those in the lowest index category and 80 percent of those in the highest index category received excellent or very good ratings. Since 36 percent of all the ratings were very good or excellent, we could predict 56 percent correctly by choice. Using this index composed of the two independent variables that had the strongest effect on the dependent variable

<p>| TABLE 25: Applicants Receiving Excellent or Very Good Ratings by Rank of Current Department: 10 Programs |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|</p>
<table>
<thead>
<tr>
<th>Program</th>
<th>Current Ranks</th>
<th>Department Ranks</th>
<th>Applicants</th>
<th>Applicants</th>
<th>Applicants</th>
<th>Applicants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics</td>
<td>0.75</td>
<td>0.95</td>
<td>0.70</td>
<td>0.80</td>
<td>0.70</td>
<td>0.80</td>
</tr>
<tr>
<td>Chemistry</td>
<td>0.65</td>
<td>0.85</td>
<td>0.60</td>
<td>0.70</td>
<td>0.60</td>
<td>0.70</td>
</tr>
<tr>
<td>Biology</td>
<td>0.55</td>
<td>0.75</td>
<td>0.50</td>
<td>0.60</td>
<td>0.50</td>
<td>0.60</td>
</tr>
<tr>
<td>Economics</td>
<td>0.45</td>
<td>0.65</td>
<td>0.40</td>
<td>0.50</td>
<td>0.40</td>
<td>0.50</td>
</tr>
<tr>
<td>Sociology</td>
<td>0.35</td>
<td>0.55</td>
<td>0.30</td>
<td>0.40</td>
<td>0.30</td>
<td>0.40</td>
</tr>
<tr>
<td>Anthropology</td>
<td>0.25</td>
<td>0.45</td>
<td>0.20</td>
<td>0.30</td>
<td>0.20</td>
<td>0.30</td>
</tr>
<tr>
<td>Geography</td>
<td>0.15</td>
<td>0.35</td>
<td>0.10</td>
<td>0.20</td>
<td>0.10</td>
<td>0.20</td>
</tr>
</tbody>
</table>

74

The first row of numbers indicates the best scores. Number of applicants in parentheses. Rank of department was broken down into quintiles using the 2 score.
rating, we were able to increase the number of cases we could predict correctly to 60 percent. This suggests still further that our ability to predict ratings from these independent variables is not greatly enhanced by constructing such indices. One reason why the overall predictability is not greater is that there are relatively few cases in these extremes. For example, only 6 percent of all the cases are in the highest index category and only 8 percent of all the cases are in the lowest. A great majority of the cases are in the middle index categories between 4 and 7, where the percentage difference is only 10 points. Since this distribution is not artifactual but is representative of the distribution of the scientists who applied to the 10 programs we studied at the nasa, it cannot be discounted. Since there is not a great deal of variance in the independent variables, they are of little use in making better predictions of the dependent variable. This is one reason why citations are not a strong predictor of ratings in algebra, fluid mechanics, anthropology, and economics.

However, the data displayed in Table 26 also allow us to compare scientists who are at different ends of the index combining citations and rank of department. Let us compare the probabilities of receiving excellent or very good ratings among scientists at the two extremes. Table 26 shows a 46 percentage point difference between the two groups. This substantial difference in probability does not contradict our findings of overall low predictability because we are dealing with only 14 percent of the total sample. Table 26 shows, as one would expect and hope, that scientists with a very high index are much more likely to receive high ratings than those with a very low index. However, the number of scientists between the two extremes is so large that the index has very little predictive value.

<table>
<thead>
<tr>
<th>TABLE 27</th>
<th>Proportion of Variance Explained (R²) on Rating by Type of Current Institution (Ph.D. or not): 10 Programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebra</td>
<td>0.01</td>
</tr>
<tr>
<td>Anthropology</td>
<td>0.01</td>
</tr>
<tr>
<td>Biochemistry</td>
<td>0.02</td>
</tr>
<tr>
<td>Chemical Dynamics</td>
<td>0.00</td>
</tr>
<tr>
<td>Ecology</td>
<td>0.04</td>
</tr>
<tr>
<td>Economics</td>
<td>0.07</td>
</tr>
<tr>
<td>Fluid Mechanics</td>
<td>0.00</td>
</tr>
<tr>
<td>Geophysics</td>
<td>0.04</td>
</tr>
<tr>
<td>Meteorology</td>
<td>0.02</td>
</tr>
</tbody>
</table>
| Solid-State Physics | 0.01 |}

The data in Table 27 distinguish applicants currently employed in Ph.D.-granting institutions from those employed elsewhere. This has virtually no influence on ratings of proposals by peer reviewers. Thus the criticism that the peer review system unfairly favors applicants from prestigious Ph.D.-granting institutions are not supported by these data.

The data on the rankings of the departments in which the applicants earned their Ph.D.'s showed the extent to which this variable was correlated with ratings given by peer reviewers. Table 28 shows that the rankings of Ph.D. departments of applicants explain very little variance in ratings received.

**PROFESSIONAL AGE AND PEER REVIEW RATINGS**

Some critics of the peer review system hold that young, inexperienced applicants have less chance to receive funds than their more experienced older colleagues. We have data on the ages of applicants and on the numbers of years since applicants received their Ph.D.'s, which we call their professional age. The results of this analysis are presented in Table 29. The findings are clear. In five of the programs professional age explains no variance in the ratings given by peer reviewers. In four programs professional age explained only 1 percent of the variance in ratings. These data strongly suggest that young people have just as good a chance to receive favorable ratings of their proposals as do their older, more experienced colleagues. This conclusion is supported by
the results reported in Table 30, which shows the influence of academic rank (only for those employed in academic institutions) of applicants on ratings received. A high correlation would indicate that applicants with high academic rank have a better chance of getting favorable ratings than applicants of lower rank. Once again, the proportions of explained variance are either nonexistent or very small. Apparently, full professors do not have a significantly better chance than their lower-ranked colleagues.

Table 31 presents the relationship between professional age and ratings received for each of the 10 programs, using tabular analysis. In algebra, ecology, and meteorology, applicants who have received their Ph.D.'s within the last 5 years have slightly higher probabilities of receiving excellent or very good ratings than do applicants who received their Ph.D.'s more than 5 years ago. In most of the other programs the difference in the proportion receiving excellent or very good ratings between relatively new Ph.D.'s and older Ph.D.'s is slight. The one program that shows a moderate relationship between these two variables is solid-state physics. Sixty-nine percent of scientists who received their Ph.D.'s more than 5 years ago received excellent or very good ratings, and 49 percent of those who received their Ph.D.'s within the last 5 years received excellent or very good ratings.

### COMBINING THE NINE CHARACTERISTICS

We conclude our analysis of the influence of principal investigators' characteristics on reviewer ratings with Table 32. This table presents the amount of variance explained in ratings by all nine characteristics of applicants, using multiple regression analysis. The table shows that the characteristics of principal investigators on whom we have data explain only a small portion of the variance in ratings in all 10 programs.

Economics is the program in which the largest proportion of variance in ratings—21 percent—is explained by the combination of nine characteristics of the principal investigators. We should point out that we do not know the extent to which even this variance in ratings is a result of the influence of these nine characteristics of applicants and how much is due to an unknown correlation between the characteristics of appli-
The fact that the nine characteristics explain so little variance in ratings is contrary to the expectations of many people. We must therefore consider carefully the implications of the findings. Clearly, we have results that, at first, appear the more obvious, the number of reviewers for a program and the number of reviews of the same proposal. The correlations are artificially reduced, the correlation of the number of reviews and the number of reviews of a program. As reported, the number of reviewers and the number of reviewers of a program were the independent variables. The results indicated that the proportion of variance of the ratings received for the various proposals. This analysis led to a greater understanding of the factors that influence the ratings received for scientific research proposals.
Influence of Characteristics of Applicants on Ratings

The table below presents the mean standardized deviation of ratings for each program, along with the mean standard deviation of ratings. The mean standardized deviation is calculated by dividing the mean deviation of ratings by the standard deviation of ratings for each program. This provides a measure of the relative importance of each characteristic in determining the ratings.

### Table 3: Mean Standard Deviation of Ratings

<table>
<thead>
<tr>
<th>Program</th>
<th>Mean Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dentistry</td>
<td>0.45</td>
</tr>
<tr>
<td>Medicine</td>
<td>0.50</td>
</tr>
<tr>
<td>Nursing</td>
<td>0.64</td>
</tr>
</tbody>
</table>

The mean standardized deviation for Dentistry is the lowest, indicating that this characteristic has the least influence on the ratings. In contrast, the mean standardized deviation for Nursing is the highest, suggesting that this characteristic has the greatest influence on the ratings.

The correlations between the mean standardized deviation and the characteristics of the applicants were also calculated. These correlations are presented in Table 4.

### Table 4: Correlation Coefficients

<table>
<thead>
<tr>
<th>Program</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dentistry</td>
<td>0.12</td>
</tr>
<tr>
<td>Medicine</td>
<td>0.32</td>
</tr>
<tr>
<td>Nursing</td>
<td>0.25</td>
</tr>
</tbody>
</table>

The correlation coefficient for Dentistry is the lowest, indicating a weaker relationship between the characteristic and the ratings. In contrast, the correlation coefficient for Medicine is the highest, suggesting a stronger relationship between the characteristic and the ratings.

### Table 5: Standard Deviation of Ratings

<table>
<thead>
<tr>
<th>Program</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dentistry</td>
<td>0.40</td>
</tr>
<tr>
<td>Medicine</td>
<td>0.50</td>
</tr>
<tr>
<td>Nursing</td>
<td>0.60</td>
</tr>
</tbody>
</table>

The standard deviation of ratings for Dentistry is the lowest, indicating less variability in the ratings. In contrast, the standard deviation of ratings for Nursing is the highest, suggesting greater variability in the ratings.

In conclusion, the characteristics of the applicants, such as academic performance and research experience, play a significant role in determining the ratings. The standardized deviation and correlation coefficients provide a quantitative measure of the influence of these characteristics on the ratings.
section are not primarily a result of low levels of agreement among the reviewers of each proposal. It is still a question needing further research to determine exactly how much reviewer disagreement exists and the significance of such disagreement for the peer review process.

This will be fully investigated and reported in Phase 2.

The data in this section show that, on the average, reviewers' numerical ratings of proposals are not heavily influenced by the characteristics of the applicants. Perhaps they are more likely to be influenced by the reviewers' perceptions of the quality of science proposed.