THE EVALUATION STUDY

WestEd conducted a multiple method study of the NSF Graduate Research Fellowship Program designed to update information on program outcomes and assess the contributions of the GRF Program to NSF goals. The last cohort of NSF fellows to be included in an analysis of outcomes was 1981 fellows, and both the science and engineering environment and the university context of graduate education have changed dramatically in the past two decades. The study included both GRF Program competitions, which award fellowships to Graduate Fellows (GF) and Minority Graduate Fellows (MGF). Additionally, Women in Engineering (WENG) award recipients were included. The purpose of the GRF Program evaluation is to provide data on program effectiveness that will be useful to ongoing NSF efforts to strengthen its programs and support the agency's commitment to make optimal use of its resources across program options that address its goals. WestEd was assisted in study design by NSF program staff and a panel of experts (Appendix A).

Previous Studies

The National Research Council (NRC) has conducted four major studies of outcomes of the GRF Program (Harmon, 1977; Snyder, 1988; Baker, 1994, 1995). Outcome indicators have included traditional measures of academic career success such as completion rates, time to degree, subsequent academic appointment, success in obtaining research grants, and, in the first study, publications and citations. These studies looked at completion rates and career plans of cohorts of new NSF fellows (1952-1972; 1967-1976; 1972-1981; and 1979-1981 cohorts respectively). They used NSF's annual Survey of Earned Doctorates (SED) to track completion rates and time to degree of fellows by gender, ethnicity, and discipline. As a comparison group, the Snyder and Baker studies used GRF applicants who were assigned to Quality Group 2[2] but who were not awarded fellowships.

All of the National Research Council studies used a measure of program quality to assess the stature of fellows' graduate programs as well as the quality of subsequent academic appointments. The measures of success and explanations for differences in outcomes for different groups relied on existing databases, especially the SED, NSF/NIH (National Institutes of Health) postdoctoral and research grant files, and the sample of NSF fellows included in the Survey of Doctorate Recipients (SDR).[3]

---

[1]GRF applicants in each competition are assigned to quality groupings during the review process. There are four quality groupings: Quality Group 1s receive fellowship awards; Quality Group 2s and Quality Group 3s receive either awards or Honorable Mentions; and Quality Group 4s do not receive either awards or Honorable Mentions. All applicants in Quality Group 1 receive fellowships based strictly on competitive ratings. However, those applicants assigned to Quality Groups 2 and 3 receive either a fellowship or an Honorable Mention based on competitive ranking plus other factors, such as field of study, level of earned graduate credits, and geographic location of applicant's high school.

[2]The Survey of Doctorate Recipients is a longitudinal survey. The primary source of the sampling frame is the SED. The target population and sample frame consist of all individuals under the age of 76 who received a research doctorate in science or engineering from a U.S. institution and were residing in the United States on April 15 in the year of administration (biennial). Recipients of research doctorates are added each time the survey is conducted, and those individuals over age 75 are dropped. A total of 50,000 individuals with research doctoral degrees in science and engineering were included in the 1995 survey. Approximately 15% of NSF fellows who complete their degrees are included in the SDR sample.

The four NRC studies confirmed that the students supported by the GRF Program are well qualified, attend outstanding graduate programs, complete their degrees in a shorter time than the comparison group of non-awardees, and are likely to become successful scientists and engineers. An "award" effect was identified because Quality Group 2 awardees achieved greater success than did Quality Group 2 non-awardees. However, such a conclusion implies comparability between these two groups that may not supported by the selection process. More recent NRC studies also reported that the large gender differences in career outcomes found in earlier studies appeared to be disappearing.

The authors of these studies acknowledged limitations in the data used in the reports and suggested additional approaches and follow-up questions. Harmon (1977) devoted an appendix to suggestions for a more comprehensive study, including collecting information to address the reasons for individuals' decisions and information on the impact of the GRF Program on the academic community. Another problem was the use of appropriate measures of success for NSF fellows and the GRF Program. The NRC reports used several traditional measures of success in academic careers such as appointments to faculty positions and publications and citations. Snyder (1988) noted the "limited nature of measures used for career outcomes"[4] and suggested that "more adequate responses could be obtained through direct follow-up with a sample of fellows and a matched sample of non-awardees" (p. 166). Baker (1995) commented that "further efforts to deepen our understanding of these disturbing statistics [low completion rates for African-American fellows] will be necessary in order to formulate appropriate policies to deal with them" (p. 28).

NSF also funded a comprehensive assessment of the effectiveness of the GRF Program that used secondary data analysis. The focus of the report (Sween, 1982) was on traditional outcome measures of academic career achievement, including degree completion, time to degree, securing academic positions, receipt of research grants, supervision of dissertations, and publications/citations. Findings confirmed the value of the GRF Program in contributing to scientific career achievements.

There were two reports prepared by NSF about the GRF Program that focused on the MGF competition and addressed an important NSF policy concern, the diversity of the SMET workforce. An internal preliminary report on the first four years of the MGF competition (Johnson, 1987) offered an overview of participants, the institutions they attended, and the results to date of their academic achievements. Of interest in this report is the finding that the first cohort (1978-1981) of MGF recipients differed significantly from their GF counterparts:

> In addition to the predictable differences in the ethnic composition of Fellows in each program, the Graduate Fellows were younger, there were proportionately more males, and a larger percentage were majoring in the natural sciences versus the social sciences. On the other hand, a larger percentage of Minority Graduate Fellows had previous graduate work at the time of application. (p. 3)

[4] Appointments to faculty positions are obtained from the National Faculty Directory, which uses data from college catalogs and questionnaires. There is a problem with coverage, as well as with identical or ambiguous names (Snyder, 1989, p. 176). The data on research support cover only NSF/National Institutes of Health/Alcohol, Drug Abuse, and Mental Health Administration awards. Because of the difficulty in using publications and citation data that are available from the Institute for Scientific Information databases, only the Harmon (1977) study used that information. And finally, no information is available on those pursuing non-academic research careers (with the exception of a small number of fellows who are included in the longitudinal Survey of Doctorate Recipients).

Legacy for Tomorrow (NSF, 1988) documented the stories of 51 MGF recipients who received their awards in 1986 or earlier and included a discussion by a panel of leaders in higher education. Two things are particularly striking about this account compared to the formal NRC studies of the GRF Program. First, the individual stories of the MGF recipients included their struggles and the importance of the NSF fellowship to their educational experiences and career choices. Second, attention was given to the context of graduate education and how the MGF awards made an impact on issues such as breaking down the barriers to graduate study in science and engineering for underrepresented groups.

In his introduction to *Legacy for Tomorrow*, NSF Director Erich Bloch wrote:

> The Minority Graduate Fellowship Program, which is highlighted in this publication, addresses the first and most crucial component of our science and engineering base–people. It is no secret that there are alarming shortages of graduate level engineers, mathematicians, and scientists. It is also no secret that certain groups remain, unfortunately and to our nation's cost, underrepresented in these ranks. (p. vi)

Although the MGF competition was discontinued in 1998, the GRF Program continues to play an important role in achieving the NSF goal of diversity in the SMET workforce.

Evaluation Questions

The purpose of the NSF Graduate Research Fellowship Program is to ensure for the nation the future stream of highly qualified scientists and engineers to undertake research and development. Inherent in this purpose is the commitment to a diverse workforce that includes participation of women and underrepresented minorities in successful careers in science and engineering. This study was designed to update findings regarding program outcomes through traditional measures such as those used in earlier studies, and provide new information on recent NSF fellows. The evaluation questions that guided this study are:

How do NSF applicants and fellows compare to non-applicants in the same academic programs, and do fellows attend institutions with the highest reputations?

- How do the academic qualifications of NSF applicants and fellows compare to non-applicants in the same academic programs?
- Are other funding opportunities more attractive to graduate students and why?
- What are the quality characteristics of institutions chosen by NSF fellows?

Do recent NSF fellows show evidence of more timely completion of degree and early career success?

- Do NSF fellows demonstrate higher degree completion rates and shorter time to degree than non-fellows do?
- What GRF Program or institutional factors contribute to success?
- Do NSF fellows show more evidence of early publication success?
- Do NSF fellows show more evidence of early research success?
- Do NSF fellows accomplish non-traditional activities that demonstrate excellence in scientific and engineering research in new ways?

Do GF and MGF recipients experience similar educational and career success?

- Do MGF recipients complete degrees at the same rates and in comparable time to other fellows, or their peers?
- Do MGF recipients experience similar early career success to other fellows, or their peers?
- Do recipients of the Women in Engineering and Women in Computer and Information Science awards have similar patterns of success compared to other fellows?

What factors contribute to differences? Does the individual award aspect of the NSF Graduate Research Fellowship enhance the educational experience and career options of fellows?

- Does receipt of an NSF fellowship change a fellow's enrollment choices?
- Does individual funding enhance the graduate experience and access to career opportunities?

Methodology

The study used multiple methods to answer the evaluation questions: secondary data analysis, administration of a survey questionnaire, and interviews from institutional site visits. The design builds on strengths of prior studies that used existing data sources but also addresses their weaknesses by employing both quantitative and qualitative data collection methods. For details of the methods employed for data collection and analysis, see Appendix B, Methodology.

Secondary Data Analysis of Attendance Patterns and Completions

We matched the Cumulative Index (CI)[5] to the Survey of Earned Doctorates from 1999 to measure attendance patterns, completion rates, and time to degree of NSF fellows who received first-year awards between 1979 and 1993. We included the 1979-1981 cohorts, even though they were also included in the two most recent studies (Baker, 1994; 1995) in order to update the completion data. Since the Minority Graduate Fellowship competition also began full implementation in 1979, starting then allows a nearly complete analysis of MGF recipients through 1993. Program quality rankings were obtained from the National Research Council's most recent study of doctoral programs (National Research Council, 1995). As in prior studies, we compared the performance of NSF fellows to non-awardees in the Quality Group 2 category. We examined differences by discipline grouping, gender, program quality, and GF and MGF competitions.

The Graduate Student Follow-up Survey

The Graduate Student Follow-up Survey (Appendix C) included items on educational background and experience, careers, and financial support during graduate school. In order to allow for comparison, the survey questionnaire contained some questions included in a recent University of California, Berkeley study that was administered to a sample of 1982-1985 Ph.D. completers in several disciplines. We also reviewed findings of the Berkeley study for the 107 NSF fellows included among the 6000 doctoral recipients surveyed (Cerny & Nerad, 1999).

By selecting cohorts of NSF fellows and program peers who *entered* their graduate programs between 1989 and 1993 we were able to gather data from respondents who did not complete graduate programs to discover reasons for not completing degrees and to gather information on their career paths. Although some respondents were still enrolled in graduate programs, most had transitioned from graduate school to postdoctoral study or employment. This enabled us to capture experiences of recent graduate students, including those who chose to leave their graduate programs without completing.

We sent the Graduate Student Follow-up Survey questionnaire to the following three samples:

1. **Disciplinary Sample**: NSF fellows (termed Disciplinary fellows in this report) and a comparison group of program peers (Disciplinary peers) in four disciplines at 16 institutions who entered the same programs from 1989 to 1993 (sample=1131)

2. **Minority Graduate Fellows Sample**: 35% of all MGF fellows from 1989-1993 (sample=200)

3. **Women in Engineering Sample**: 50% of all WENG recipients from 1990-1993 (sample=143) (WENG began in 1990. Since 1994, additional awards have also been made to women in Computer and Information Sciences. We did not include computer science recipients because these awards began after our survey cohort period, 1989-1993.)

[5]The Cumulative Index of NSF Fellowship Applicants and Awardees contains information on all applications to the NSF GRF Program, including applicant demographics, educational data, test scores, and fellowship status.

The analysis was conducted by sample. Although we found differences between NSF fellows and program peers and differences across disciplines, these differences did not attain statistical significance. However, findings suggest areas for further exploration of the differences reported here. Care should be taken when looking at findings across samples because of the significant disciplinary and institutional differences in the samples and because there were some respondents included in more than one sample.

The Disciplinary Sample

The Snyder (1988) and Baker (1994, 1995) studies as well as our secondary data analysis used Quality Group 2 non-awardees as the comparison group. The strength of this approach is that those in the comparison group went through the same fellowship review process. Having done so, the Quality Group 2 non-awardees were considered comparable to fellows not only using such measures as Graduate Record Examination (GRE) scores but also with regard to their ability to write a strong research essay and the strength of personal recommendations.

There are also weaknesses to this approach. This comparison group is limited to individuals who applied for an NSF graduate fellowship, and it does not reflect a random assignment based on selection criteria. Furthermore, Quality Group 2 non-awardees are not necessarily enrolled in the same graduate programs as NSF fellows, allowing institutional effects to intrude into the design. Finally, GRF applicant eligibility requirements also influence this non-awardee comparison group since many SMET students are ineligible, including international students and those who already hold advanced degrees.

We identified a database that allowed the selection of a comparison group that is not limited to NSF applicants. Since 1989, the American Association of Universities (AAU) and the Association of Graduate Schools (AGS) have been working with the Educational Testing Service (ETS) on a database on doctoral students. The resulting AAU/AGS Doctoral Education Database contains data on doctoral students entering graduate programs since 1989, including graduate students from 40 institutions in four NSF disciplines: Biochemistry, Economics, Mathematics, and Mechanical Engineering. The AAU/AGS database includes GRE scores and undergraduate institution, but it does not contain personal recommendations or other materials that are taken into account in the awarding of GRF fellowships.

We used the AAU/AGS database to identify program peers entering the same graduate programs as NSF fellows in the same years. Anticipating a lower response rate from non-NSF fellows, we administered the Graduate Student Follow-up Survey to all NSF fellows (GF and MGF recipients) and a sample of two program peers for each fellow (2X sample) (sample=1131) in the disciplines of Biochemistry, Economics, Mathematics, and Mechanical Engineering at 16 institutions (Appendix D). The pool for the "program peers" comparison group included all students who began the same graduate programs at the same time as the NSF fellows but who did not receive (and may not have applied for) an NSF Graduate Research Fellowship.[6]

[6]The 16 institutions included 15 participants in the AAU/AGS database that was used to sample the program peers and one non-participating institution with a large number of NSF fellows that agreed to provide comparable data and a program peer sample. With this exception, we only included those institutions that participated in the database and had at least two NSF fellows entering programs in the four disciplines from 1989-1993.

Locating survey recipients proved to be a very difficult task. NSF does not track former fellows, and although most of the 16 institutions provided mailing addresses to us, many of these were no longer valid. We sent two post card follow ups to individuals whose questionnaires were not returned completed or marked undeliverable. We hired a private investigation firm to further search for current addresses, but some individuals still could not be found. More than three-quarters of those we could not locate in this sample were program peers. Discounting the individuals we could not contact (194 or 17.15%), our response rate for the Disciplinary Sample was 41.41%. Completed surveys received from 200 NSF fellows (55.71%) and 188 program peers (32.53%) have been included in the analysis for this report.

MGF Sample

We were concerned that the Disciplinary sample would not include sufficient numbers of fellows from the Minority Graduate Fellowship competition to provide us with good data on the experience of MGF recipients. To increase responses from NSF fellows from underrepresented groups, we administered the same Graduate Student Follow-up Survey to 200 MGF recipients (35% sample), regardless of discipline or institution enrolled in. The MGF sample was randomly drawn from the Cumulative Index for 1989-1993. There is no comparison group for the MGF sample.

The MGF sample included fellows in 33 disciplines at 62 institutions. Each institution was asked to provide a current mailing address, but as with the Disciplinary sample, many addresses were no longer valid. We employed the same follow-up and search procedures, and discounting individuals we could not contact (25 or 12.50%), the response rate was 49.71%. Questionnaires were received from 87 MGF recipients. The MGF sample was analyzed independently of the Disciplinary sample.

WENG Sample

In 1990, the Directorate of Engineering began providing funding to the GRF Program for additional fellowships for women pursuing graduate education in engineering. These additional fellowships were awarded in both the GF and MGF competitions. In order to see how the WENG recipients have fared, we also administered the Graduate Student Follow-up Survey to 143 WENG recipients from 1990-1993 (50% sample). The WENG sample was randomly drawn from the Cumulative Index for 1990-1993 and included NSF fellows in various sub-fields of engineering at 46 institutions–not just Mechanical Engineering, the field used for the Disciplinary sample. There is no comparison group for the WENG sample.

Each institution provided addresses for WENG fellows. After using the same follow-up and search procedures and discounting individuals we could not contact (18 or 12.58%), the response rate was 68.00%. Questionnaires were received from 85 WENG recipients. The WENG sample was analyzed independently of the Disciplinary sample.

Institutional Site Visits

We conducted site visits at six major research universities that will be identified as Institutions A-F.[7] Four of these universities enroll very large numbers of NSF fellows, and the other two were selected for institutional and geographical balance. The six institutions included two private and four public universities located in the Northeast (1), Southeast (1), Midwest (1), and on the West Coast (3). The two private institutions enroll more graduate students than undergraduate students, and the six universities range in total enrollment from 10,000 to 40,000. In the academic year of the site visits (1998-1999), these universities individually enrolled between 38 and 293 NSF fellows in all programs and, taken together, enrolled 962 NSF fellows. Five of the six universities we visited were also included in the survey for the Disciplinary sample. At the six institutions, we interviewed 75 administrators, faculty, and staff and 149 students (73 NSF fellows and 76 peers) in 19 departments that correspond to the four disciplines in the Disciplinary sample. Interview protocols may be found in Appendix E.

What Did We Find?

The NSF Graduate Research Fellowship Program continues to fund highly qualified graduate students and is considered to be a very prestigious award. However, the 1998-1999 stipend ($15,000/year for 3 years) and the cost of education allowance ($9,500/year for 3 years) levels are considered too low for prevailing costs. Other fellowship programs are more attractive because they offer higher stipends, expense allowances, full tuition reimbursement, more years of support, and/or networking opportunities. Unlike many of these other graduate fellowship programs, the GRF is much larger, involves a national competition, and supports more students for graduate study in a greater variety of fields. The GRF is most highly valued for its prestige and the choices it affords fellows. The GRF Program is also a major source of graduate student funding for some programs at top research universities.

A high percentage of NSF fellows complete the Ph.D. However, we found that despite qualifications and financial support, 32% of 1979 to 1983 NSF fellows and 27% of 1984 to 1988 NSF fellows did not complete the Ph.D. within 11 years of entering their graduate programs. Furthermore, a gap in completion rates remained between GF and MGF recipients, although this gap appears to have narrowed. We also found that 11-year completion rates for women are approaching those of men. WENG fellows' 6-year completion rates are lower than Ph.D. completion rates for all other fellows (both men and women) in Engineering.

The Graduate Education Follow-up Survey and institutional site visits gave us the opportunity to probe the reasons for choices made by students as well as to discover institutional and program factors that influenced decisions. We found a highly complicated world of graduate education populated by real people coping with the challenges and changes of the late 20th century. Costs of graduate education have grown dramatically, and graduate students are more diverse, reflecting major changes in American society. When NSF and the Graduate Research Fellowship Program were created nearly 50 years ago, it was literally a different world.

---

[7] In this report, site visit institutions are referred to as Institution A through F. References to other institutions will read Institution X. To reference quotations from site visit reports, disciplines are abbreviated as follows: BIO for Biochemistry, EC for Economics, M for Mathematics, and ME for Mechanical Engineering.

Higher education in the United States is a world of hierarchy, dominated by those institutions that receive the most funding for research and have faculty with reputations for high levels of research productivity. Among those institutions previously classified Research University 1 or 2 by The Carnegie Foundation for the Advancement of Teaching (Carnegie, 1994), there is also a reputational hierarchy. The National Research Council (NRC) ranks individual graduate programs in research universities using reputational measures (NRC, 1995).

Competition among top ranked programs for the best graduate students is fierce. Equally important, however, is the competition for admission to top programs because the prestige of graduate program and advisor are key determinants of initial appointment and academic career success (Clark, 1987; Smith, Wolf & Busenberg, 1996). Within institutions, there is a very high degree of decentralization that is particularly profound at the graduate level. Disciplines are the major determinants of the graduate experience, and there is significant variation within many disciplines (Bowen & Rudenstine, 1992). Disciplines change over time, subdividing but sometimes reorganizing around core themes. This has been true, for example, of Biochemistry.

Graduate education plays a critical role in the preparation of the academic, business, industry, government, and non-profit workforce in the United States. In 1996, for all science and engineering disciplines (excluding professional degrees), 28,458 doctorates and 131,436 master's degrees were awarded (NSF, 1999b). The nation's research and development funds support the work of both faculty and students. Over $13 billion of Federal funds went to universities and colleges for research and development in 1997 (NSF, 1998). NSF has been a major source of support for graduate education at universities that produce the research scientists of the future. NSF funding includes research grants to faculty that support graduate students who participate in funded research projects and graduate training programs.

Through the GRF, NSF supports graduate students in specified fields at institutions of their choice. NSF fellows study in a variety of disciplines, so their experiences and the impacts of the fellowship differ. We found this to be true for participants in the Graduate Student Follow-up Survey and NSF fellows interviewed during the institutional site visits. Graduate education in the United States is part of an interlocking system of higher education that includes complex dynamics, variation, and also key leveraging points. This report focuses on three transition points: entry, graduate experience, and career/life choices.