

# 3145-NEW, Evaluation of National Science Foundation's East Asia and Pacific Summer Institutes and International Research Fellowship Program

## Section B

### Statistical Methods

#### B.1. Respondent Universe and Sampling Methods

The universe of respondents for which the clearance is sought includes eight groups: (1) EAPSI fellows (n=1,434); (2) IRFP fellows (n=567); (3) EAPSI foreign hosts (up to 1,434); (4) IRFP foreign hosts (up to 1,434); (5) EAPSI US advisors (up to 1,434); (6) EAPSI unfunded applicants (n=1,401); (7) IRFP unfunded applicants (n=1,502); (8) key staff at EAPSI foreign location (n=20). Note, the number of foreign hosts and advisors may be lower than the total number of fellows if foreign hosts or advisors worked with more than one fellow. We propose to survey the universe of each of these respondent groups.

*EAPSI and IRFP Fellows.* The fellows sample includes participants from cohorts that span the program years 1999-2009 for EAPSI and 1992-2009 for IRFP.

*EAPSI and IRFP hosts.* The hosts sample includes participants from cohorts that span the program years of 1999-2009 for EAPSI and 1992-2009 for IRFP.

*EAPSI US advisors.* The universe of US advisors includes participants from cohorts that span the program years of 1999-2009 for EAPSI. IRFP fellows' advisors are not included in the survey as in contrast to EAPSI fellows, IRFP fellows have earned their degrees and therefore do not have an advisor (doctoral degree is a condition for participation).

*Unfunded applicants - comparison groups.* The evaluation design incorporates the use of unfunded applicants as a comparison group for each program. These individuals who had applied to the EAPSI and IRFP programs, but were ultimately not awarded the fellowships (referred to as "unfunded applicants"). The key advantage of using this group as a comparison is in the similarity of interest and motivation to engage in international collaboration, and intent to conduct research in foreign countries between the applicants and the fellows.

*EAPSI foreign location key staff.* Interviews will be conducted with key contact staff at each of the EAPSI foreign locations. These are representatives of the foreign partner organizations who are familiar with, and help administer, the EAPSI fellows in their countries.

We anticipate a response rate of at least 75 percent from the respondent group based on previous surveys conducted with students and early career scientists of NSF-sponsored programs. Response rates are projected based on similar surveys conducted

with samples of graduate students and early career researchers who participated in NSF programs. Table B.1 illustrates the response rates for various evaluation studies of NSF programs that surveyed graduate students and early career individuals, which were used to estimate the expected response rates for this project.

<b>Table B.1</b>		
<b>Program</b>	<b>Response Rate</b>	<b>Length of Time Between Participation and Data Collection</b>
CAREER Fellows	84%	0-10 years
IGERT Former Students	74%	0-10 years
GK-12 Fellows	MS 45% PhD 57%	5-10 years
GK-12 Fellows	MS 83% PhD 92%	0-5 years

### **B.2. Information Collection Procedures/Limitations of the Study**

The following steps will be taken to collect survey data on various populations described in the previous section.

*Step 1: mining NSF data.* NSF program data consists of e-Jackets (for years from 2001 to 2009), paper applications (for years 2000 and prior), and internal program files. Internal program files contain information useful in locating respondents (including names, discipline, institution at the time of application, address and phone number at time of application).

*Step 2: locating respondents.* Once NSF data are organized into a central database, the following steps for locating respondents will be taken:

1. Use names and other available information from the NSF records to conduct Google and other web-based searches.
  - Obtain contact information from individuals' own web pages (as academic researchers and many graduate students have research team/lab or home page websites) and verify
  - Obtain contact information from posted articles, presentations, and other materials.
2. Contact PhD advisors to request information about their students' whereabouts. This approach would probably be most effective for relatively recent applicants (within the past 5 years).
  - If the name of the advisor is missing, use ProQuest to locate the fellows dissertation abstract and document the name of the advisor (only for doctoral thesis).
3. For difficult cases for whom email addresses could not be found based on the searches in 1-2 above, use the following procedure:

- Use names and NSF program contact information to search AccurInt (database linked to LexisNexis) to verify or update the addresses and phone numbers in the NSF records.
- Mail an invitation to participate in the survey (with information on how to access the survey web site) to the latest known address. Use the US mail Forwarding Service to obtain information on the change of address. The invitation will be mailed up to three times using the forwarding address provided by the US postal service. Also, dial the phone numbers identified in AccurInt. Once individuals have been contacted, verify that the individuals are indeed the individuals being sought by the study, and then invite them to participate in the survey and provide instructions on how to access the web site.

*Step 3: Web survey.* Once approval is obtained from OMB, we will program the surveys for online data collection. The study team will test each survey system to ensure functionality and accuracy of data capture; survey data collection is scheduled to begin in fall 2010.

All subjects will be sent an invitation email by NSF, introducing the study and the contractor conducting the study (Abt Associates). Abt will follow up with another email, containing a link to the survey, username, and password. Three email reminders and three telephone reminders will be used to boost response rates. The survey will be open for two months during the academic year. Throughout the data collection cycle, a toll-free number and e-mail address will be available to ensure that potential respondents can easily and quickly obtain answers to questions or concerns.

#### **Estimation Procedure**

The purpose of this proposed activity is to collect data from participants and unfunded applicants of the EAPSI and IRFP programs in an effort to measure the initial and potential long-term impact of these programs. Analysis will include a descriptive reporting using the measures of central tendency and frequency distributions. Data from awardees and those who applied for but did not receive the fellowship awards of the EAPSI and IRFP programs will be compared using propensity score matching. Propensity score matching that will allow a comparison of the fellows (treatment group) to unfunded applicants (comparison group) selected based on their similarity to the awarded applicants. With this approach, the fellows would be compared to unfunded applicants who are as similar as possible to them in terms of observable characteristics, allowing us to determine what the fellows' outcomes would have been had they not received the IRFP or the EAPSI award, had other characteristics been equivalent. The PSM models are a way of matching members of different groups based on a range of characteristics that will allow more accurate estimates of the program effects. Appendix B contains additional details about the PSM approach.

#### **B.3. Methods for Maximizing the Response Rate and Addressing Issues of Nonresponse**

Method to maximize response rate are described in detail in section B.2. Briefly, these will include the following procedures:

1. Extensive location techniques to identify correct email address
2. Web format of the survey
3. Minimization of spam filtering
4. Invitation from NSF to participate in the study
5. Skip patterns, to reduce burden on respondents
6. Extensive email and telephone follow-up
7. Availability of a toll-free number and email address for questions.

We will examine the bias in estimates because of nonresponse by following the three steps described below. Based on the analysis we will adjust the weights of responding students to account for student nonresponse.

*1. Examination of Response Rates.* The first step will be to monitor the overall response rate, as well as by year and by relevant subgroups (e.g., by discipline, or by gender and race/ethnicity). High response rates (over 80 percent) for the entire sample as well as for subgroups might indicate no need for further analysis of bias due to nonresponse. Large differences in the response rates by strata and for subgroups serve as indicators that potential biases may exist. For example, if response rate from an important subgroup is very low then any difference in the characteristic of interest between this subgroup and other subgroups would result in a bias in the estimates. From the survey results we will examine whether there are differences in the characteristics in the subgroups, especially in a stratum where the response rate is low.

*2. Comparison of estimates based on respondents to estimates from external sources.* For questions where there are data available from an external source for some characteristic of interest (e.g disciplines, proportion n tenure track positions), we will compare the estimates from our survey responses to those from nationally available data. A large difference may indicate bias in the survey estimates assuming that the external source provides an unbiased estimate.

*3. Nonresponse Propensity Model.* Finally, should the response rate fall below 80 percent we will construct a propensity model to estimate the probability of a student in responding to the survey both for responding and nonresponding students; this is called a propensity score. The estimated propensity scores come from a logistic regression model. The model will be based on variables which are available both for nonresponding and responding students. Students will be grouped using the estimated propensity scores. Within each group we will compare the frame characteristics of responding and nonresponding students. This grouping in addition to assessing the bias will also provide a method of forming weighting classes for adjusting the weights of responding students to reduce the bias due to nonresponse.

#### B.4. Tests of Procedures or Methods

Experts in the field are part of the Advisory Panel and they have reviewed the study design and data collection instruments. The survey was pilot-tested with former graduate students, who were asked to comment on the clarity and content on the questions and to record the time required to complete the survey. Minor revisions, including shortening the length of the survey, resulted from this feedback. (The median time to completion was 32 minutes. The survey was shortened by removing individual items and sections in order to reduce the respondent burden to 30 minutes.)

Once the survey instruments are programmed, they will be tested online by Abt researchers familiar with the project.

#### B.5. Names and Telephone Numbers of Individuals Consulted

Key personnel who have been involved in the statistical aspects and who will be involved in collecting and analyzing data are presented in the table below. The contractor for collection and analysis of data in this study is Abt Associates Inc., Cambridge, MA. Staff have knowledge of statistical methods, experience in evaluation of research programs, and expertise in scientific research were involved in the design. Members of the Advisory Panel were also consulted in the design, and who may also be consulted in the analysis of data. Finally, NSF program staff members familiar with the programs have been included in the design of the evaluation.

<b>Table B.5 Individuals Consulted</b>		
<b>Name</b>	<b>Role</b>	<b>Phone</b>
<b>Abt Associates Inc.</b>		
Alina Martinez	Project Director, Senior Associate	617-349-2312
W. Carter Epstein	Associate	617-349-2543
Fatih Unlu	Economist, Scientist	
K.P. Srinath	Statistician, Survey Sampling and Methodology	301-634-1836
Luba Katz	Associate	617-349-2313
<b>National Science Foundation</b>		
John Tsapogas	Program Coordinator, Office of International Science and Engineering	703-292-7799
Jong-on Hahm	EAPSI Program Manager, Office of International Science and Engineering	703-292-7223
Susan Parris	IRFP Program Manager, Office of International Science and Engineering	703-292-7225
Edward Murdy	Senior Program Manager, Office of International Science and Engineering	703-292-8711
<b>Advisory Panel</b>		
Irwin Feller	Professor Emeritus of Economics, Penn State	814-865-0691
Susan Cozzens	Professor of Public Policy and Director of its Technology Policy and Assessment Center,	404-385-0397

	Georgia Institute of Technology	
Terrence Russell	Principal, Terrence Russell LLC, Executive Director Emeritus, the Association for Institutional Research	850-228-9273
Christopher Hill	Director, Doctoral Program in Public Policy, George Mason University	703-993-2270
Nicholas Vonortas	Dept of Economics; Director, Center for International Science and Technology Policy, George Washington University	202-378-6230

## **Appendices**

Appendix A: Survey Instruments

Appendix B: PSM Details