Supporting Statement (3145-0187)

Evaluation of the Graduate Teaching Fellows in K-12 Education (GK-12) Program: OMB 3145-0187

Section B

Introduction

B.1. Respondent Universe and Sampling Methods

Survey Samples

The survey data collection process will include a sample of participants from the universe of GK-12 projects funded between 2000 and 2004 (138 projects). Starting with a single sampling frame of awards we will identify five samples: GK-12 Fellows, Comparison graduate students, PI's, Fellows faculty advisors, and teachers. In developing a sampling frame, we have considered how to maximize analytical strength while minimizing the implementation costs and burden to participants. We have described each sample in more detail below.

GK-12 Fellows Sample:
A stratified sample of Fellows will be chosen. The target population for this survey is all Fellows who are Masters and PhD Fellows who participated sometime between the 1999-2000 and 2005-2006 academic years. There are 2,435 Fellows in this population.
For the selection of the sample of Fellows, we will stratify the population of Fellows defined above using two stratification variables each with two categories. This results in four strata for sample selection. The table below shows the strata and the distribution of the population of Fellows by strata. We will select a sample of Fellows within each stratum that will enable comparisons between Masters' Fellows at institutions with High and Low research intensity[1] and between PhD level Fellows at institutions with High and Low research intensity.

Distribution of the Population of Fellows by Strata

	Master's	PhD	All
High Research Intensity	395	816	1,211
Low Research Intensity	812	412	1,224
All	1,207	1,228	2,435

---

[1] Institutions were divided into High and Low Research Intensity categories using the Carnegie basic classification. High Research Intensity= RU/VH: Research Universities (very high research activity) and Low Research Intensity= Assoc/Pub2in4: Associate's--Public 2-year colleges under 4-year universities; RU/H: Research Universities (high research activity); DRU: Doctoral/Research Universities; Master's L: Master's Colleges and Universities (larger programs); and Spec/Med: Special Focus Institutions--Medical schools and medical centers.

As indicated above, one of the objectives of the study is to compare subgroups. We want the sample size to be large enough to be able to detect differences between groups relating to important characteristics with 80% power when we do a two-sided statistical test at 5% level of significance. The sample sizes also depend on the cost of collection of data and the available budget. We give in the following table the required sample sizes for detecting differences between population percentages with 80% power. The differences are in percentage points. We have taken into account the population size in each group while computing the sample sizes. The finite population size has also been taken into account for determining the sample sizes since the sample sizes are large in comparison to our population size.

Required Sample Size for Detecting Various Differences between Master's fellows at High Research Intensity Institutions and Master's fellows at Low Research Intensity Institutions

Size of Difference Between Two Groups	Master's fellows at High Research Intensity Institutions	Master's fellows at Low Research Intensity Institutions
8 percentage points	240	348
9 percentage points	217	302
10 percentage points	196	262

Required Sample Size for Detecting Various Differences between Ph.D. fellows at High Research Intensity Institutions and Ph.D. fellows at Low Research Intensity Institutions

Size of Difference Between Two Groups	Ph.D. fellows at High Research Intensity Institutions	Ph.D. fellows at Low Research Intensity Institutions
8 percentage points	349	248
9 percentage points	302	222
10 percentage points	263	200

In order to detect differences of 8 percentage points between the two groups with 80% power and assuming a 75% response rate and an 80% find rate, we propose a sample of 1,738 Fellows. We show the allocation of the total sample of Fellows to each stratum in the table below.

Distribution of the Total Sample by Strata

	Master's	PhD	All
High Research Intensity	362	503	865

Low Research Intensity	503	370	873
All	865	873	1738

We propose to draw a systematic sample of Fellows within each stratum. For example, 503 Fellows will be selected from the 812 available in the stratum which is Master's-low research.

Comparison graduate students:
For each Fellow included in the Fellows Sample with a degree, we will select from the same institution a comparison non-GK-12 graduate student who is a United States citizen from the same discipline, and matches the Fellow as closely as possible on the following variables, in priority order:
1. Degree level (Masters, PhD)
2. Faculty advisor
3. Year of enrollment in degree program (plus or minus 1 year)
4. Enrollment status (still enrolled/ graduated)

A slightly larger number of Non-GK-12 graduates (1,862) will be selected relative to GK-12 Ph.D. recipients to control for their anticipated lower response rate of 70%.

Distribution of the Total Sample by Strata

	Master's	**PhD**	**All**
High Research Intensity	388	539	927
Low Research Intensity	539	396	935
All	927	935	1862

PI Sample:
We will survey the Principal Investigator (PI) or Project Coordinator of every award from the selected cohorts (2000-2004). This gives us a sample of 138 PIs.

Fellows' faculty advisor Sample:
For each Fellow included in the Fellows Sample, we will select the faculty member who advised the Fellow during his or her first year of GK-12 participation.

K-12 Teacher Sample:
A stratified systematic sample of teachers will be selected. All teachers, from the 138 awards, who have participated between September, 2005 and May, 2007 will be included in the sampling frame from which the teacher sample will be drawn. The teacher sample

will include all types of teachers, both those who worked directly and indirectly with Fellows. There are 3,383 teachers in this population.

We will stratify the population of teachers defined above using two stratification variables (school level and level of involvement with a fellow) resulting in six strata for sample selection. The table below shows the strata and the distribution of the population of teachers by strata. The number of teachers sampled from each stratum will be proportional to the population number of teachers within each stratum.

Distribution of the Population of Teachers by Strata

Involvement with Fellow	Elementary	Middle	Secondary	All
Direct	681	703	585	1969
Indirect	472	469	473	1414
All	1153	1172	1058	3383

As indicated above, one of the objectives of the study is to compare subgroups. We want the sample size to be large enough to be able to detect differences between groups relating to important characteristics with 80% power when we do a two-sided statistical test at 5% level of significance. Assuming a 75% response rate and a 50% find rate (because of the high transience rate associated with the teacher population, we anticipate that some proportion of teachers selected will not be locatable), we propose a sample of 1,867 Teachers. This will allow us to detect differences of 11 percentage points (based on the number of teachers available in each stratum in the population and our assumptions that this is the smallest percentage point difference we can hope to detect) between the two groups with 80% power. We show the allocation of the total sample of teachers to each stratum in the table below.

Total Sample of Teachers by Strata

	Elementary	Middle	Secondary	All
Direct	376[=141/(0.5*0.75)]	387	323	1085
Indirect	261	259	261	781
All	637	645	584	1867

The following chart provides estimates of the sizes of the various universes that will be sampled.

Summary Table of the Five Samples for Survey Data Collection:

Population	Universe Size	Sample Size
GK-12 Fellows	2435	1738
Comparison graduate students	Unknown	1862
K-12 teachers	3383	1867
GK-12 principal	138	138

investigators		
Faculty advisors	2435	1738

Interview Samples

The interview data collection process will include conducting interviews with samples of GK-12 Fellows, K-12 teachers, GK-12 PIs, and participating faculty members. We have determined the number of respondents we would like to survey for each of the groups. However, we plan to use the data from our proposed survey instruments to inform our sampling methods for selecting our interview samples.

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

Internet-based surveys will be used to collect data from GK-12 Fellows, a comparison group of non-GK-12 graduate students, PIs, faculty advisors, and K-12 teachers to determine the impacts of the program on the participants. Any conclusions drawn from this may be biased, as there is no way to control who is participating in these programs. It is possible that both the character of the program and the outcomes for participants are more the result of their inherent tendency to seek the GK-12 experience than they are the effect of NSF funding.

We are not including a comparison group for teachers. Any comparison teachers would need to be drawn from participating GK-12 schools, as surveying teachers from non GK-12 schools lies outside the current scope of this project. There are serious concerns for drawing a comparison group of non-GK-12 teachers from within GK-12 schools including potentially significant differences in interest and aptitude for STEM education; the risk of contamination; and the low number of appropriate comparison teachers within middle and high schools (where there are often only one or two teachers per subject area).

B.2.1. Statistical Methodology for Stratification and Sample Selection

Survey Samples

Fellows Sample:
As mentioned in section B.1. (Respondent Universe and Sampling Methods), we will stratify the population of Fellows and a comparison group of non GK-12 Fellows using two stratification variables each with two categories. The variables are the following: degree level (e.g., Masters vs. Phd) and research intensity of the university (e.g., high vs. low intensity) which results in four strata for sample selection. We chose to stratify our sample along these variables because we predict that they will have a strong correlation with the proposed Fellows outcomes that we are measuring in this study.
For the selection of the sample within each stratum Fellows will be sorted by the following variables:
  5. Estimated graduation status of each Fellow (Enrolled; Graduated)

6. Cohort
7. Award
8. STEM discipline of each Fellow

After sorting, a systematic sample of Fellows will be selected within each stratum. For example, 503 Fellows will be selected from the 812 available in the stratum which is low research and Master's. Similar independent selections will be made in other strata. Selecting Fellows systematically after sorting by the variables specified above will ensure that representation in the sample for various characteristics used in sorting will be in the same proportion as in the population. If some awards have more Fellows than other awards, then the sample also will have more Fellows from that award than other awards. This proportional representation will result in more precise estimates of program outcomes, as compared to a simple random sample.

Teachers Sample:
For the selection of the sample within each stratum Teachers will be sorted by the following variables:
1. Cohort
2. Last Year of Participation
3. Number of Years of participation

After sorting, a systematic sample of Teachers will be selected within each stratum. For example, 376 Teachers will be selected from the 681 available in the stratum that is elementary direct. Similar independent selections will be made in other strata. Selecting Teachers systematically after sorting by the variables specified above will ensure that representation in the sample for various characteristics used in sorting will be in the same proportion as in the population. If some cohorts have more Teachers than other cohorts, then the sample also will have more Teachers from these cohorts than other cohorts. This proportional representation will result in more precise estimates of program outcomes, as compared to a simple random sample.

See section B.1 (Respondent Universe and Sampling Methods) for further details on stratification and sampling methods.

Interview Samples

As mentioned before, we plan to use the data from our proposed survey instruments to inform our sampling methods for selecting our interview samples.

B.2.2. Estimation Procedure

The purpose of this proposed activity is to collect data from recently graduated Fellows and compare their short term career outcomes with recent STEM graduates who did not participate in the GK-12 program to measure the long-term impact of the GK-12 program on Fellows. Data will be collected from Fellows, participating teachers, students, PIs, faculty advisors, Institutions of Higher Education, and K-12 schools to assess the overall

impacts of the program on its many participants. Analysis will begin with a descriptive analysis of the survey data and move on to other types of analysis as appropriate.

B.2.3. Degree of Accuracy Needed for the Purpose Described in the Justification

Not Applicable

B.2.4. Unusual Problems Requiring Specialized Sampling Procedures

Not Applicable

B.2.5. Use of Periodic (Less Frequent Than Annual) Data Collection Cycles

Not Applicable

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

In an effort to increase overall survey response rate, follow-up with respondents will be multi-modal. Respondents will initially be sent an email containing a link to an Internet survey. The emails will contain an individualized link for each respondent that they can click on and that will take them directly to the survey. Respondents to Internet surveys will have the option of pausing survey completion and returning at a later time to finish. Telephone and email follow-up will be used for non-respondents.

B.4. Tests of Procedures or Methods

A GK-12 Planning Session with program participants and experts, as well as a thorough review of a selected sample of annual and final project reports, informed the development of the survey instrument and interview protocol. The survey instruments and interview protocols developed for this data collection will be pilot-tested through cognitive interviews in Spring 2008.

B.5. Names and Telephone Numbers of Individuals Consulted

Agency Unit
Carol Stoel, National Science Foundation, 703-292-8624
William Neufeld National Science Foundation, 703-292-5148

Contractor
Abt Associates Inc.
4550 Montgomery Ave, Suite 800 North,
Bethesda MD 20854