

## B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

### B.1 Respondent Universe and Sampling Procedure

The GSS is an annual census of all eligible institutions. The universe is intended to cover all U.S. academic institutions that offer graduate (master’s and PhD or equivalent) degree-credit programs in science and engineering, as defined by NSF, and graduate programs in health fields, as defined by NIH.

#### B.1.1 Discussion of Institutional Frame

An institution is considered eligible for the GSS if it meets the following criteria:

- Grants at least one master’s or doctoral degree in at least one program listed in a GSS-eligible field (See Attachment 5 for the list of GSS fields).

Following extensive research in 2010, a number of potentially eligible institutions were surveyed in the 2011 GSS cycle. A total of 165 new schools were added in 2011 (see Exhibit 6). The new schools tend to have fewer eligible units and the relative proportion of part-time students is higher than in the “core” GSS institutions. About 39% of new institutions have only one eligible program and only 7% of the new institutions have more than 5 GSS eligible degree programs. After further review of the data collected in the 2012 cycle, 162 of the “new” institutions were finally found to be eligible for the survey. Therefore, the total of 727 institutions which consisted of 848 schools and 14,429 units were surveyed in 2012.

**Exhibit 6. Number of Core and New GSS Institutions, 2011-2012**

Year	Institutions	Schools	Organizational units			Graduate enrollment		
			Total	Master's	Doctorate	Total	Full time	Part time
All Institutions								
2011	730	852	14,217	4,704	9,513	653,170	466,017	187,153
2012	727	848	14,429	4,766	9,663	659,282	469,054	190,228
Core								
2011	565	686	13,785	4,330	9,455	626,820	457,292	169,528
2012	565	684	13,952	4,362	9,590	627,243	459,498	167,745
New								
2011	165	166	432	374	58	26,350	8,725	17,625
2012	162	164	477	404	73	32,039	9,556	22,483

The data collected from the newly eligible GSS institutions were not included in the 2011-2013 survey data published by NSF because of the ongoing evaluation of data reliability for the new schools, and of the efforts to coordinate the taxonomy of disciplines across all the surveys conducted by NCSES (see Section B.1.2). Since many of the newly eligible institutions have only a single eligible program, there was a concern that the upcoming taxonomy change in SEH fields may result in changes to institutional eligibility. The revised GSS eligible SEH fields based on the final NCSES' taxonomy of disciplines will be used to determine the institution eligibility beginning in 2014.

In the 2015 and 2016 cycles, the GSS will undergo a population review to identify any new institutions that should be added to the list of eligible institutions. This review will be conducted in a manner consistent with earlier reviews to ensure that all eligible (in-scope) schools are surveyed. The population review is conducted by consulting the following sources to identify potentially eligible schools that are not currently included in the GSS frame:

- Carnegie Classification—Master's colleges and universities I and II, doctoral/research universities extensive and intensive, and specialized institutions (<http://classifications.carnegiefoundation.org/>)
- NCES IPEDS database—All 4-year institutions that offer programs in engineering, engineering technologies/technicians, biological and biomedical sciences, mathematics and statistics, physical sciences, psychology, social sciences or health professions, and related clinical sciences (<http://nces.ed.gov/collegenavigator/>); a crosswalk between CIP codes and GSS codes is provided in Attachment 6
- CGS membership directory (<http://www.cgsnet.org/current-cgs-members/>)
- Association of American Medical Colleges list of accredited U.S. MD-granting medical schools and similar membership lists (<https://members.aamc.org/eweb/DynamicPage.aspx?site=AAMC&webcode=AAMCOrgSearchResult&orgtype=Medical%20School/>)
- National Postdoctoral Association (<http://www.nationalpostdoc.org/>)
- Association of American Universities list of member institutions (<http://www.aau.edu/about/article.aspx?id=5476/>)
- Higher Education Directory
- NSF Higher Education Research and Development Survey
- NSF Survey of Science Engineering and Research Facilities

### **B.1.2 NCSES Taxonomy of Disciplines and Changes to GSS Eligible Fields**

During the past year, NCSES has conducted a thorough review of the Taxonomy of Disciplines (TOD) which is used across the Center's surveys. As a result, some SEH fields were reclassified. Four codes that were previously eligible for GSS as SEH were changed to non-S&E fields in the TOD, and consequently will be deleted from the GSS beginning in 2014. Below is a list of those fields and the number of units in GSS 2012 that were assigned the codes:

- GSS Code 913 (Public administration) – 210 units
- GSS Code 920 (Family/consumer sciences) – 111 units
- GSS Code 930 (Communications) – 262 units
- GSS Code 940 (Architecture) – 55 units

The modification to align with the TOD will result in a deletion of 638 units from the 2013 GSS. In addition, during the review process an additional 17 units were found to be ineligible based on results of reconciliation to the CIP-GSS crosswalk. Thus, an estimated total of 655 units will be deleted prior to the start of the 2014 GSS.

### **B.2 Description of Survey Methodology and Statistical Procedures**

The GSS is a Web-based survey. In the fall 2012 survey, 93% of coordinators provided data via the Web and 7% provided their data in part or completely via data files. Hard copy worksheets are provided to the new SCs to allow them to see the types of information requested in the survey. Occasionally, institutions provide data in an Excel file or another medium which the contractor reformats for entry into the web instrument, especially during the missing or inconsistent data follow-up phase.

Each institution has one or more coordinators that manage data collection activities. Some institutions have separate coordinators for the graduate enrollment section and the postdoc section, and some have separate coordinators for the graduate and medical schools. For new institutions, NSF mails the president a letter of invitation and asks the president to name a coordinator or coordinators for the survey and to verify the institutions' eligibility. Institutions that do not respond to the letter are followed up via phone call and e-mail.

The GSS instrument comprises two parts. In Part 1, the coordinator updates a list of all eligible units (departments, programs, research centers, and health care facilities) in the school and classifies each unit by its GSS code (field). A crosswalk between GSS codes and CIP codes is provided to respondents (see Attachment 6 for the crosswalk used in the 2013 GSS). For established GSS schools, this activity involves verifying the eligibility of units pre-populated from the previous survey round, confirming GSS codes, adding any newly eligible units, and deleting defunct units. All Part 1 activities are completed by the coordinator.

In Part 2, data for each unit are entered or uploaded by the coordinator, or by designated unit respondents (URs), whom the coordinator may assign as needed. Part 2 requests details about graduate students, postdocs, and NFRs in each GSS-eligible unit. A paper worksheet corresponding to Part 2 is provided with the survey materials sent to each coordinator, if requested (See Attachment 1). Information can be compiled on this worksheet in preparation for entry into the Web survey. Once a UR has completed Part 2, the UR submits data to the coordinator, who reviews and revises the data as needed. Once all units are ready to submit, the coordinator submits the school's data to the GSS survey contractor. After data submission, the coordinator can only view their data. The data are then reviewed by the survey contractor and any questionable items are flagged for data review and follow-up. If the coordinator needs to make a revision, the survey contractor can roll back the data submission so that the coordinator can make the needed change and resubmit prior to the final survey close-out date.

The SC serves as the point of contact at the institution for all internal and external communications about the GSS. The SC's responsibilities include notifying the URs of their assignments and ensuring that the UR submits the completed data by the established due date.

Each GSS survey cycle begins with a pre-data collection e-mail to the SC from the previous survey cycle to determine if he/she is still the appropriate contact for the upcoming cycle. The e-mail is typically sent in early September with a telephone follow-up if confirmation is not received. Once the SC is confirmed/updated, data collection commences. Data collection begins in October with an e-mail and FedEx package providing the SC with Web access information and information about the GSS-eligible degree programs.

Data collection procedures for the 2014-17 survey cycles are expected to be similar to those used in the 2013 GSS. The data collection plan is included in Attachment 7.

### B.2.1 Imputation for Item Nonresponse in the GSS

Imputation is used when missing data are present for any item. The 2012 GSS collected responses for 355 items related to four categories of graduate students (part- and full-time) and personnel (postdocs and NFRs). All missing data were imputed. The imputation rates for these variables ranged from 0.9% to 7.6%, with a mean of 5.0%. The imputation procedures used for the 2014-2016 GSS will remain similar to those used in the GSS for 2012 and 2013. A simplified summary of the imputation methods used in 2012 follows.

The imputation procedure used for a given question for a given unit depended on whether data were provided in any prior survey cycle and whether totals were provided in the current cycle. The method used under each of four conditions is shown in Exhibit 7.

**Exhibit 7. Imputation Methods Used by Condition for 2012 GSS**

	<b>Current Survey Cycle Totals Available</b>	<b>No Current Survey Cycle Totals Available</b>
<b>Prior Survey Cycle Data Available</b>	1. Carry forward (details only)	2. Carry forward (totals and details)
<b>Prior Survey Cycle Data Unavailable</b>	3. Nearest neighbor (details only)	4. Adjusted Enrollment for graduate student totals; Nearest neighbor for other totals and all details

When the 2012 total was reported without complete data for details, but the details were reported by the unit in a previous survey cycle, the details were imputed using a carry-forward (CF) method, whereby the prior year’s distribution of the total over the details was applied to the 2012 total.

When details needed to be imputed but a prior year’s data were not available, a nearest neighbor (NN) was identified from the set of units that responded in the 2012 GSS. When graduate student details were being imputed, the NN selected had full-time and part-time graduate enrollments that were most similar to the imputee’s enrollments.<sup>1</sup> When postdocs and NFR details were being imputed, the total number of postdocs and NFRs were used to choose the NN. If the total number of postdocs or NFRs was unknown, the total numbers of full-time and part-time

---

<sup>1</sup> In cases where the unit to be imputed had provided only total enrollment (full-time and part-time combined), the total was split into full-time and part-time enrollment before imputation using the CF or NN method.

students were used to find proper donors. In either case, the details were imputed by distributing the total according to the nearest neighbor's distribution.

When data is missing for an item in 2012, total imputation by a CF method was employed if data from a prior survey cycle was available. First, the total was imputed by multiplying the prior year's total by an inflation factor to account for year-to-year change. The details were then imputed by applying the prior year's distribution to the imputed total. The same procedure was used in the 2011 imputations.

In rare instances where neither current year totals nor data from a prior year were available, a method called adjusted enrollment (AE) was used for imputation of graduate student data. Unlike the CF and NN methods, which use only GSS data, the AE method uses IPEDS data to estimate the graduate student totals by gender. In this method, for each gender category, the institutional graduate enrollment totals were obtained from the IPEDS Fall Enrollment survey. These totals were then distributed respectively to the totals of missing and nonmissing units according to the IPEDS distributions over the CIP codes in the IPEDS completion survey within gender category by following a crosswalk between the GSS Code and CIP codes. (See Attachment 6.) If there were multiple GSS codes matched with one CIP code in the same institution, the total for all missing units was evenly distributed to each of the missing units. These totals were further distributed to detailed cells using the NN method.

Since the IPEDS data do not include counts of postdocs or NFRs, the GSS required a different method when these data were missing and no prior data were available. The unit's full-time and part-time graduate student enrollment figures, as reported or imputed for the 2012 GSS, were used to identify a NN donor from the pool of GSS units. The donor's postdoc and NFR data were then used to impute the missing data.

There are exceptions to these procedures. Some institutions report counts at the institution level or school level without allocating the counts to the individual units. For these special cases, the institution or school totals are allocated to the units according to historical proportions, and the unit totals are allocated to the details according to the methods described above.

The 2012 GSS survey frame contains 14,429 units. Of the 14,429 eligible organizational units for the 2012 GSS, a total of 12,315 units were full respondents (85.3%), 2,023 units were partial respondents (14.0%), and 91 units were total non-respondents (0.6%) for which key totals

and details were imputed for all graduate students, postdocs, and NFRs data. Exhibit 8 summarizes the number of units imputed for the 4 key totals (total FTs, total PTs, total postdocs, and total NFRs) by each imputation method. Over 98 percent of full-time and part-time graduate student key totals did not require imputation. Key totals for postdocs and NFRs required slightly more imputation, where 3.9 and 7.4 percent of totals needed imputation, respectively. Among the key totals for postdocs and NFRs, the CF method was the most frequent imputation method used for key totals, followed by NN. Less than 0.2 percent of the cases required special imputation procedures.

**Exhibit 8. Imputation Methods for 2012 GSS Key Totals, Counts and Percentage of Total Cases**

Imputation Method	Graduate Student Full-time		Graduate Student Part-time		Postdoc		NFR	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Total	14,429	100	14,429	100	14,429	100	14,429	100
No imputation	14,302	99.1	14,244	98.8	13,872	96.1	13,367	92.6
Carry Forward	76	0.5	133	0.9	401	2.8	814	5.6
Nearest Neighbor	1	0.0	2	0.0	130	0.9	233	1.6
Adjusted Enrollment	1	0.0	1	0.0	0	0.0	0	0.0
Special Case	28	0.2	23	0.2	15	0.1	15	0.1

Source: 2012 GSS.

**B.3 Methods Used To Maximize Response Rate**

Because the GSS is designed to produce estimates for all academic institutions that offer graduate degree programs in SEH fields, care is made to maximize response rates. The GSS contractor staff work closely with the SCs to build strong working relationship with all participating institutions and try to ensure that all contacts are positive.

Survey techniques proven successful in past surveys will again be used to maximize the GSS response rate. These techniques include

- Early pre-data collection confirmation of the SC
- Two-part GSS data collection to ensure early notification of unit respondents of their assignments. The first part entails a review/update by the coordinator of the GSS-relevant programs and notification of unit respondents to begin their data reporting assignments. The second part is the unit-level reporting of counts of graduate students, postdocs, and NFRs

- Separate due dates for each of the two GSS parts to help identify at the earliest juncture those institutions that might be potential nonrespondents
- Targeted e-mails and telephone follow-up based on response status
- Availability of knowledgeable GSS contractor staff to provide assistance to the coordinators and unit respondents
- Multiple modes of data collection allowed (Web, data file uploads, etc.)
- GSS Help desk staff available to respond to telephone and e-mail questions and concerns raised by institution staff
- Presentations at AIR and CGS meetings demonstrating the Web-based data collection system and discussing any proposed changes
- The inclusion of cover letters explaining how the provided data are used
- The inclusion in the survey package and in the GSS Web survey of a “crosswalk” listing the fields of study for which data are requested for the GSS along with the NCES CIP codes for these fields as published in *A Classification of Instructional Programs*. This crosswalk is for the convenience of those institutions using CIP codes in reporting their enrollment and degrees to the IPEDS system (See Attachment 6)
- Enlistment of others at the institution, as appropriate, to gain cooperation

These methods have proven successful in the past, as evidenced by response rates.

Exhibit 9 displays unit, school, and institutional response rates for the 2010-2012 survey cycles.

**Exhibit 9. Institution, School, and Unit Response Rates: 2010–12**

	Complete respondents			Partial respondents			Nonrespondents		
	2010	2011	2012	2010	2011	2012	2010	2011	2012
Institution	98.3% <i>n</i> =564	96.8% <i>n</i> =707	97.2% <i>n</i> =707	1.0% <i>n</i> =6	0.8% <i>n</i> =6	0.1% <i>n</i> =1	0.7% <i>n</i> =4	2.3% <i>n</i> =17	2.6% <i>n</i> =19
School	98.3% <i>n</i> =680	97.2% <i>n</i> =828	97.6% <i>n</i> =828	1.0% <i>n</i> =7	0.6% <i>n</i> =5	0.1% <i>n</i> =1	0.7% <i>n</i> =5	2.2% <i>n</i> =19	2.2% <i>n</i> =19
Unit	85.4% <i>n</i> =11,703	84.8% <i>n</i> =12,503	85.3% <i>n</i> =12,315	13.7% <i>n</i> =1,880	14.0% <i>n</i> =1,987	14.0% <i>n</i> =2,023	0.9% <i>n</i> =128	1.2% <i>n</i> =177	0.6% <i>n</i> =91

During the next three cycles, more concerted efforts will be made to encourage and help the participating institutions to move toward greater use of the data upload feature. This feature will reduce the respondent burden dramatically in schools with large numbers of units by preparing data files directly from their administrative databases and submitting those files via the GSS web system. In May 2013, the NSF presented a session at the AIR Annual Forum demonstrating the GSS data upload features and solicited feedback from the participants, many of

whom are GSS respondents. As a result, a “test” website was created and the SCs were invited to use to test their data upload programs prior to the launch of the 2013 GSS.

Efforts will be made to identify institutions with large numbers of units, and educate them of the benefits of using the GSS data upload feature. The procedures that demonstrate the steps involved in downloading files from the institutional systems and aggregating the information to be uploaded will be prepared and shared with institutions that might benefit from this approach.

#### **B.4 Testing of Procedures**

NSF has sponsored methodological research for every survey cycle to help improve the GSS survey. In this section, methodological investigations taken since the last OMB clearance submission in 2011 are reviewed.

##### **B.4.1 Examination of Impact of Methodological Changes to Observed GSS Trends**

No major changes were made to the GSS questionnaire during the past three years. However, based on the results of a 2009 Postdoc Pilot Study and a Recordkeeping Survey conducted with GSS respondents, a new series of items about postdocs and other doctoral nonfaculty researchers (NFRs) was included in the 2010 GSS questionnaire to improve the reporting of postdocs and NFRs. During the past several years, we have examined how these changes have impacted the data and GSS trends.

##### *Expanded Postdoc Items and Introduction of Postdoc Coordinator Role*

The total number of postdocs reported in the 2010 GSS increased to 63,415 in 2010, an increase of 10% over the 2009 total of 57,805 postdocs and 25% over the 2007 total of 50,840 postdocs. These 1- and 3-year growth rates are the highest in the history of the GSS and likely reflect improved reporting in addition to the continued expansion of postdoc employment in academia.

To ensure that GSS data users understand the impact of the methodological changes, an InfoBrief was written, which highlighted how the changes contributed to the higher number of postdocs reported in 2010 (<http://www.nsf.gov/statistics/infbrief/nsf13334/>). The results of the analyses suggest that asking institutions to appoint a GSS postdoc coordinator was instrumental in the improved reporting; new postdoc coordinators identified additional units with postdocs, as

well as more postdocs within existing units. The results of this examination suggest that the postdoc data collected beginning in 2010 are the most accurate and comprehensive to date and that aggregate trends by discipline and demographics were largely unaffected by recent changes in reporting.

#### *Expanded Non-Faculty Doctorate Researcher (NFR) Items*

In 2010, items were also added to the NFR section of the GSS questionnaire. The importance of these new items, including their use in the Carnegie Classifications, was communicated to the institution Presidents as well as school and postdoc coordinators. In part as a result of changes to the questionnaire and procedures, the number of NFRs reported by GSS institutions increased dramatically from 14,059 in 2009 to 21,145 in 2010, and has remained relatively steady since that time<sup>2</sup>. While the analysis of these changes is not yet complete (we recently revised imputation procedures and imputed the detailed 2010 -2012 NFR data), it appears that the percentage of schools reporting no NFRs in two or more years is much lower in 2010-2012 than in 2007-2009, 47.6% compared to 63.9%, as shown in Exhibit 10. It also appears that consistency of NFR reporting has increased since the changes were introduced, with more schools reporting similar NFR counts in across all three years. While schools are not as consistent in their reporting of NFRs as they are in reporting postdocs and graduate students, there has been an improvement over prior years of GSS.

---

<sup>2</sup> The analyses of NFR data is limited to core schools, so that comparisons across years can be made.

### Exhibit10. Consistency of NFR, Postdoc, and FT Grad Student Counts

	Level of consistency							
	NFRs 2010-2012		NFRs 2007-2009		Postdocs 2010-2012		FT grad student 2010-2012	
	n	%	n	%	n	%	n	%
Total core schools with data in all 3 years	678	100.0	674	100.0	678	100.0	678	100.0
Schools with zeroes in 2 or more years	323	47.6	431	63.9	245	36.1	28	4.1
Similar counts in all 3 years*	158	23.3	68	10.1	255	37.6	347	51.2
Similar counts in 2 of 3 years*	75	11.1	41	6.1	113	16.7	205	30.2
Different counts in all 3 years	122	18.0	134	19.9	65	9.6	98	14.5

\*Similar count is defined as counts less than 25 or a percent change less than 15.

A short web-based debriefing survey was also conducted in 2013 with SCs coordinators concerning NFR reporting, following a memo to OMB requesting clearance for methodological research. The debriefing survey revealed that the majority of the respondents understood the NFR questions. However, some reporting errors were identified that involved duplicate reporting of postdocs and NFRs in some institutions. This information was used to identify this pattern among all GSS schools and correct the misreported data. The research report on the NFR data is under development.

#### B.4.2 Anticipated Methodological Research

*Research on the Postdoc/NFR “unknown” response categories.* Use of the “unknown” category by institutional respondents and ways to reduce its use will be examined, as well as the possibility of imputing responses in the “unknown” category. We are currently reviewing the reporting of the “unknown” responses in a few Postdoc and NFR data items, especially the item that asks institutions to report on origin of the doctorate degree of the NFRs. During the 2013 cycle, we identified schools that had some units using the “unknown” category for this data item and other units reporting actual counts for the same data within the same school. This information was shared with SCs to find out the data availability in the institution’s administrative system.

### B.4.3 Changes to the 2014 GSS

No major changes to the 2014 are anticipated, other than the revisions to the taxonomy described in Section B.1.2. Attachment 9 contains a listing of the minor changes to be made to the 2014 GSS, based on the methodological work and the experience gained in conducting prior survey rounds.

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

The individuals consulted on GSS technical and statistical issues are listed in Exhibit 11.

#### Exhibit 11. Individuals Consulted on GSS Technical and Statistical Issues

Name	Affiliation	Telephone number
Ms. Kelly Kang GSS Survey Manager	National Science Foundation, NCSES, Arlington, VA	703-292-7796
Ms. Emilda Rivers HRS Program Director	National Science Foundation, NCSES, Arlington, VA	703-292-7773
Ms. Wan-Ying Chang Mathematical Statistician	National Science Foundation, NCSES, Arlington, VA	703 292-2310
Ms. Rebecca Morrison Survey Statistician	National Science Foundation, NCSES, Arlington, VA	703 292-7794
Dr. Patricia Green Project Director	RTI International Chicago, IL	312-456-5260
Mr. Peter Einaudi Data Analysis Task Leader	RTI International Research Triangle Park, NC	919-541-8765
Ms. Jamie Friedman Data Collection Task Leader	RTI International Chicago, IL	312-456-5262
Dr. Jean Lennon FFRDC Postdoc Survey Task Leader	RTI International Research Triangle Park, NC	919-485-2654
Mr. Jim Rogers Data Delivery Task Leader	RTI International Research Triangle Park, NC	919-541-7291
Mr. Bob Steele Systems Development Task Leader	RTI International Research Triangle Park, NC	919-316-3836
Dr. Rachel Harter Mathematical Statistical Task Leader	RTI International Research Triangle Park, NC	919-541-6472