

2017–2019
Survey of Graduate Students
and
Postdoctorates in Science
and Engineering

OMB Supporting Statement
Section B

September 2017

TABLE OF CONTENTS

Section
n

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS.....	1
B.1 Universe and Sampling Procedure.....	1
B.1.1 Discussion of Institutional Frame.....	1
B.1.2 Separate Reporting of Master’s and Doctoral Student Data.....	2
B.1.3 NCSES Taxonomy of Disciplines and Changes to GSS Eligible Fields.....	2
B.1.4 Collection of Data Based on CIP Codes.....	3
B.2 Survey Methodology.....	4
B.2.1 Data Collection.....	5
B.2.2 Re-classification of GSS Codes in the Taxonomy of Disciplines (TOD).....	6
B.2.3 Imputation for Item Nonresponse.....	6
B.3 Methods Used to Maximize Response Rate.....	9
B.4 Testing of Procedures.....	10
B.5 Individuals Consulted.....	11

EXHIBITS

Number

Exhibit 8. Number of GSS Institutions, Schools, Units, and Enrollment, 2014-2016.....	3
Exhibit 9. Number and Percent of Units Affected by Newly Ineligible Fields.....	3
Exhibit 10. Imputation Methods Used by Condition for 2016 GSS.....	7
Exhibit 11. Imputation Methods for 2016 GSS.....	9
Exhibit 12. GSS Institution, School, and Unit Response Rates: 2014–16.....	10
Exhibit 13. Individuals Consulted on GSS Technical and Statistical Issues.....	11

LIST OF ATTACHMENTS

- Attachment 1. GSS Instrument Screen Shots
- Attachment 2. Federally Funded Research and Development Centers (FFRDCs) Postdoc Survey Instrument Screen Shots
- Attachment 3. First Federal Register Notice
- Attachment 4. GSS Schedule
- Attachment 5. GSS Code List—Complete List of GSS Eligible Fields and Codes
- Attachment 6. GSS Data Collection Plan
- Attachment 7. GSS Data File Upload Templates
- Attachment 8. CIP and GSS Code Crosswalk
- Attachment 9. GSS Worksheet
- Attachment 10. List of Changes to GSS Taxonomy
- Attachment 11. GSS Taxonomy Tool Prototype
- Attachment 12. GSS Coordinator Survey Report Executive Summary

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

B.1 Universe and Sampling Procedure

The GSS is an annual census of eligible institutions. The GSS universe is intended to cover all academic institutions in the United States and its territories that grant research-oriented master's degrees or doctorates, appoint postdocs, or employ non-faculty researchers (NFRs) in science, engineering, and health (SEH) fields as of the fall term. An institution is considered eligible for the GSS if it 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).

B.1.1 Discussion of Institutional Frame

In 2014, the survey frame was updated following a comprehensive frame evaluation study. The study identified potentially eligible but not previously surveyed academic institutions in the United States with research-oriented master's or doctorate-granting SEH programs. A total of 151 newly eligible institutions were added, and two private for-profit institutions offering mostly practitioner-based graduate degrees were determined to be ineligible. See Exhibit 8 for a comparison of the number of GSS institutions, schools, units, and enrollment in 2014-2016. For more information on the changes to the frame and the impact of the frame changes, see *Assessing the Impact of Frame Changes on Trend Data from the Survey of Graduate Students and Postdoctorates in Science and Engineering*¹. In the 2016 GSS cycle, to be consistent with the NCSES Higher Education Research and Development (HERD) survey, for-profit institutions were declared ineligible since they are typically not research institutions but rather focus on practitioner-oriented degrees.

¹ Arbeit CA, Einaudi P, Green P, Kang KH. 2016. Assessing the Impact of Frame Changes on Trend Data from the Survey of Graduate Students and Postdoctorates in Science and Engineering. Special Report NSF 16-314 (<https://nsf.gov/statistics/2016/nsf16314/>).

Exhibit 8. Number of GSS Institutions, Schools, Units, and Enrollment, 2014-2016

Year	Institutions	Schools	Units			Graduate Enrollment		
			Total	Master's	Doctorate	Total	Full-time	Part-time
All institutions								
2014	706	821	14,845	4,820	10,025	666,586	492,170	174,416
2015	711	824	15,202	4,947	10,255	685,397	506,262	179,135
2016	714	828	15,853	5,115	10,738	684,825	508,773	176,052
Doctorate institutions								
2014	406	521	13,140	3,115	10,025	588,952	452,801	136,151
2015	412	525	13,506	3,251	10,025	604,944	464,695	140,249
2016	415	529	14,188	3,450	10,738	609,420	468,678	140,742
Master's institutions								
2014	300	300	1,705	1,705	na	77,634	39,369	38,265
2015	299	299	1,696	1,696	na	80,453	41,567	38,886
2016	299	299	1,665	1,665	na	75,405	40,095	35,310

na = not applicable.

B.1.2 Separate Reporting of Master's and Doctoral Student Data

In past survey cycles, the GSS collected aggregated data on graduate students that could not be parsed by degree level. The 2017 GSS will introduce the separate reporting of enrollment and financial support data for master's and doctoral students. Increasingly over time, the GSS data users have been requesting separate data for master's and doctoral students because the field concentration and funding patterns are different between the two group. More granular data enables greater precision in understanding the educational pipeline for the SEH labor force.

B.1.3 NCSES Taxonomy of Disciplines and Changes to GSS Eligible Fields

Starting in the 2017 survey cycle, the GSS taxonomy is aligning with the Taxonomy of Disciplines (TOD) adopted by NCSES to standardize the disciplinary fields reported by all of its surveys. This alignment will result in the following changes in GSS:

- The number of broad fields will be reduced from 15 to 13. Two broad fields— Communication, and Family and consumer sciences and human sciences— will become ineligible; one broad field, Natural resources and conservation sciences will be split from Agricultural Sciences as a new broad field; and Neurobiology and neuroscience will become subfields under Biology and biomedical sciences.

- Subfields will be reorganized, rendering some subfields ineligible, and leading to the addition of several newly eligible subfields. Ten detailed fields will be deleted,² nine detailed fields will be added,³ and six detailed fields previously reported under other fields will be collected separately.⁴ In addition, some fields will have name changes to better align with the CIP program titles. For example, Biological Sciences will become Biological and Biomedical sciences; Earth, atmospheric, and ocean sciences will become Geoscience, atmospheric, and ocean sciences; and Electrical engineering will change to Electrical, electronics, and communications engineering.
- Some fields will become ineligible. The newly ineligible fields are Architecture, Communications, and Public administration. Family and consumer sciences and human sciences will also become ineligible except for the Human development subfields which will be moved under the Social sciences field. See Exhibit 9 for the number and percent of GSS units affected by these changes. These changes will lead to approximately 5.7% fewer graduate students counted in the GSS.

Exhibit 9. Number and Percent of Units Affected by Newly Ineligible Fields

Newly Ineligible Fields	Current Total	New Total	N Change	Percent Change
All Units	15,202	14,691	-690	-4.5
Fully Ineligible				
Architecture	54	0	-54	-100.0
Public Administration	232	0	-232	-100.0
Communication	279	0	-279	-100.0
Partially Ineligible (estimated)				
Nutrition	136	69	-67	-49.1
Family and Consumer Sciences	123	65	-58	-47.2

B.1.4 Collection of Data Based on CIP Codes

In the 2017 survey cycle, the GSS will collect disciplinary field data from institutions based on the CIP codes rather than NCSES's GSS codes. The collection of data by CIP codes can potentially reduce response burden because these codes are commonly used at institutions.

² The 10 detailed fields are: Anatomy 601; Biometry and epidemiology 604; Cell and molecular biology 607; Ecology 608; Entomology and parasitology 609; Sociology and anthropology 909; Public administration 913; Family and consumer sciences and human sciences 920; Communication 930; and Architecture 940.

³ The 9 new detailed fields are: Environmental science and studies 510; Forestry, natural resources, and conservation 511; Biostatistics and bioinformatics 618; Cell, cellular biology, and anatomical sciences 619; Ecology and population biology 620; Epidemiology 621; Molecular biology 622; Biomedical sciences 623; and Human development 915.

⁴ The 6 detailed fields that are currently reported under other detailed fields are Biological and biosystems engineering 115; Nanotechnology 116; Materials sciences 205; Criminal justice - safety studies 911; International relations and national security studies 912; and Public policy analysis 914.

CIP is the academic field taxonomy used by the NCES for the IPEDS, a mandatory reporting requirement for institutions receiving Title IV funding. The results of the GSS Coordinator Survey, conducted in July 2016, indicated respondents' ability to provide GSS data using CIP codes for graduate student demographic and financial support information. The schools will have the option of using either CIP codes or GSS codes for reporting postdoc and NFR data only. To help the transition to GSS schools of using CIP codes, schools will be allowed to report the graduate student data in 2017.

Using CIP codes to collect GSS data will also ease the respondent burden associated with the implementation of the new NCSES TOD. The TOD is designed to improve consistency of disciplinary fields between the NCSES surveys and the CIP. However, as discussed above, implementing the TOD in the GSS will require reclassifying units at institutions to the revised GSS codes. Collecting data using CIP codes will allow automatic recoding of the units to the new NCSES TOD in the GSS web instrument, instead of manual recoding by survey coordinators (SCs), thereby reducing the burden associated with the taxonomy change.

B.2 Survey Methodology

The GSS has been a Web-based survey for over ten years. Each institution has one or more SCs 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. Each GSS survey cycle begins with a pre-data collection e-mail to the previous survey cycle's SC 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.

For new institutions, NSF mails the president a survey invitation letter that asks the president to name a SC for the survey and to verify the institutions' eligibility for GSS. Institutions that do not respond to the letter are followed up via phone call and e-mail. Hard copy GSS worksheets are provided to the new institutions to allow them to see the types of information requested in the survey.

The SC serves as the point of contact at the institution for all internal and external communications about the GSS. The SC may choose to delegate some reporting activity to unit respondents (URs) at their institutions or they may report the GSS data themselves. If using URs, 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. The 2017 data collection plan, including a timetable and communications with GSS stakeholders, is included in Attachment 6.

The past GSS web data collection comprises two parts. In Part 1, the SC updates a list of all eligible units in the school and classifies each unit by its GSS code (field). For established GSS schools, this activity involves verifying the eligibility of units pre-populated from the previous year, adding any newly eligible units, and deleting defunct units. All Part 1 activities are completed by the SC.

In Part 2, data for each unit are provided by the SC or the URs. Part 2 data collection requests detailed information about graduate students, postdocs, and NFRs in each unit. The SC submits the Part 2 data to NSF once data for all units are completed. After the data submission, the SC can only view their data. The data are then reviewed; any questionable items are flagged for data review and follow-up with the SC as necessary. If the SC needs to make a revision, the web access to their data submission is restored so that the SC can make the needed changes and resubmit data prior to the final survey close-out date.

B.2.1 Data Collection

In the 2017 survey cycle, the SCs will be asked to prepare data files that can be uploaded directly into the GSS web survey instrument for the units that enroll graduate students, and/or employ postdocs or NFRs. The SCs will be provided with survey variable and file specifications for each type of GSS data requested—graduate students, postdocs, and NFRs—as well as file templates (in the form of Microsoft Excel spreadsheets; see Attachment 7) to organize their data. Two options will be provided for uploading their data—these options are described in Section A.3.

The SCs who are not ready or unable to provide data through data upload method will be allowed to 'opt-out' in the the 2017 survey cycle and provide their data through the manual entry of requested data in a series of grids on the GSS web instrument. A hard copy of the GSS worksheet that corresponds to the GSS web instrument will be provided with the survey

materials to the SCs, if requested (See Attachment 9). Information can be compiled on this worksheet for each unit prior to data entry into the web instrument.

B.2.2 Re-classification of GSS Codes in the Taxonomy of Disciplines (TOD)

As noted above, the GSS field taxonomy will be revised to align to the new NCSES' TOD starting in 2017. GSS codes will still be used for data reporting purposes, and implementing the TOD-aligned GSS codes will require reclassification of the GSS fields impacted by the taxonomy change. The GSS recoding process will vary depending on SCs' use of data uploads and availability of CIP codes associated with the units.

For SCs uploading data with CIP codes, the recoding of GSS codes will be completed automatically as part of the data upload process, without further input from respondents. This should mitigate reporting burden for SCs using this method. For SCs who are unable to upload data, or have units that do not have associated CIP codes (e.g., research centers), the SC will need to recode the units associated with GSS codes impacted by the taxonomy change. To assist SCs, they will be provided with a revised GSS Code List, a GSS/CIP Crosswalk and a list of GSS codes with changes. Examples of these materials are provided in Attachments 5, 8, and 10, respectively.

SCs that have units without associated CIP codes will need to use a Taxonomy Tool provided in the GSS web instrument to identify the impacted units and recode them to revised GSS codes before either uploading or reporting the data for those units. The Taxonomy Tool is expected to be used in the 2017 survey cycle as all the units are recoded to reflect the change in GSS Taxonomy in the first year. After the 2017 survey cycle, the burden associated with the taxonomy change is expected to be minimal. A prototype of the Taxonomy Tool is provided in Attachment 11.

B.2.3 Imputation for Item Nonresponse

Imputation is used for item nonresponse. The 2016 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 6.3%, with a mean imputation rate across all items of 3.8%. The imputation procedures in 2017 will remain similar to those used in the past, although they may be modified to accommodate changes in the availability of prior data due to the separate reporting of master's

and doctoral students. A simplified summary of the imputation methods used in recent GSS cycles 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 10.

Exhibit 10. Imputation Methods Used by Condition for 2016 GSS

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

When the 2016 total was reported without complete detailed data, but the details were reported by the unit in a previous survey cycle, the details were imputed using a carry-forward (CF) method. Under the CF method, the prior year’s distribution of the total over the details was applied to the 2016 total.

When the 2016 total was reported without complete detailed data, but a prior year’s data were not available, the details were imputed using a nearest neighbor (NN) method. The NN selected for the imputation was dependent on the type of detail requiring imputation (e.g., graduate enrollments, postdocs, or NFRs details). In all cases, the details were imputed by distributing the total according to the nearest neighbor’s distribution.

When data was missing for an item in 2016, 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.

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 and CIP codes (see Attachment 8). 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 2016 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 2016 GSS survey frame contains 15,853 units. Of the 15,853 eligible organizational units for 2016, a total of 13,617 (85.9%) units were classified as complete respondents, 2,157 (13.6%) units were partial respondents, and 79 (0.5%) units were total non-respondents for which key totals and details were imputed for all graduate students, postdocs, and NFRs data. Exhibit 11 summarizes the number of units imputed for the 4 key totals (total full-time graduate students, part-time graduate students, total postdocs, and total NFRs) by each imputation method. Over 99 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 2.5 and 5.5 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 11. Imputation Methods for 2016 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	15,853	100.0	15,853	100.0	15,853	100.0	15,853	100.0
No Imputation	15,718	99.1	15,703	99.1	15,464	97.5	14,988	94.5
Carry Forward	93	0.6	108	0.7	318	2.0	752	4.7
Nearest Neighbor	0	0.0	2	0.0	62	0.4	104	0.7
Adjusted Enrollment	4	0.0	4	0.0	3	0.0	3	0.0
Special Case	32	0.2	30	0.2	0	0.1	0	0.0

B.3 Methods Used to Maximize Response Rate

Because the GSS is designed to produce estimates for all U.S. postsecondary institutions that offer graduate degree programs in SEH fields, care is made to maximize response rates and thus reduce the likelihood of biased estimates. The survey staff work closely with the SCs to build strong working relationships 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 and their contact information
- Targeted e-mails and telephone follow-up based on response status
- Availability of knowledgeable survey staff and GSS Help Desk provide respond to questions, concerns and assistance to the SCs and unit respondents via telephone and emails.
- Multiple modes of data collection allowed (web instrument, two data upload options)
- The inclusion of cover letters explaining how the provided data are used
- The inclusion of a “crosswalk” listing the fields of study for which data are requested along with CIP codes for these fields. This crosswalk is for the convenience of the institutions using CIP codes in reporting their enrollment and degree award data to the NCES’ IPEDS data collection (see Attachment 8)
- Enlistment of others at the institution, as appropriate, to gain cooperation

These methods have proven successful in the past, as evidenced by high response rates. Exhibit 12 displays unit, school, and institutional response rates for the 2014-2016 survey cycles.

Exhibit 12. GSS Institution, School, and Unit Response Rates: 2014–16

	Complete Respondents			Partial Respondents			Nonrespondents		
	2014 ^a	2015	2016	2014 ^a	2015	2016	2014 ^a	2015	2016
Institution	98.7% n=697	98.0% n=697	98.0% n=700	0.3% n=2	0.6% n=4	0.4% n=3	1.0% n=7	1.4% n=10	1.5% n=11
School	98.9% n=812	98.3% n=810	98.3% n=814	0.2% n=2	0.4% n=3	0.4% n=3	0.9% n=7	1.3% n=11	1.3% n=11
Unit	86.4% n=12,832	83.6% n=12,714	85.9% n=13,617	13.2% n=1,966	15.8% n=2,405	13.6% n=2,157	0.3% n=47	0.5% n=83	0.5% n=79

^a In previous reports, these data were referred to as “2014new.” They include newly-eligible institutions that were added as a result of the 2014 frame evaluation study and do not include schools deemed ineligible as a result of that study.

In addition to the methods listed above, a series of workshops and presentations were offered at professional conferences that GSS SCs frequently attend to introduce institutions to the changes planned as part of the 2017 data collection. To date, presentations and workshops have been offered at the Association for Institutional Research Annual Forum (AIR), regional AIR conferences, and the Council on Graduate Schools. Additionally, GSS will conduct a series of webinars to provide specific instructions and support for SCs navigating the changes in GSS data collection.

B.4 Testing of Procedures

NSF has sponsored methodological research for every survey cycle to improve the survey. With the changes being made to the 2017 GSS, NCSES anticipates several methodological studies to determine the efficacy of the revised data collection and reporting procedures. Among the activities envisioned over the next three years:

- Site visits to selected institutions to explore issues related to GSS response burden, ability to provide specific data elements, and the ability to provide data based on CIP codes instead of GSS codes.
- Exploration of the feasibility of alternative methods of collecting and reporting financial support data on graduate students and postdoctoral researchers through the use of user group meetings, site visits, and pilot testing of revised data elements.
- Impact study of revised Taxonomy of Disciplines on longitudinal data trends and GSS data quality.

- A special report highlighting the split masters and doctoral data. This report would serve to examine the utility of the new GSS data and highlight the availability of these data for users.

NCSES will submit plans informing OMB of its intention to use methodological research burden hours before any applicable studies are undertaken.

B.5 Individuals Consulted

The individuals consulted on GSS technical and statistical issues are listed in Exhibit 13.

Exhibit 13. Individuals Consulted on GSS Technical and Statistical Issues

Name	Affiliation	Telephone Number
Mr. Michael Yamaner Current GSS Project Officer	National Science Foundation, NCSES, Arlington, VA	703-292-7815
Ms. Kelly H. Kang Former GSS Project Officer	National Science Foundation, NCSES, Arlington, VA	703-292-7796
Mr. John Finamore HRS Program Director	National Science Foundation, NCSES, Arlington, VA	703-292-2258
Dr. Wan-Ying Chang Mathematical Statistician	National Science Foundation, NCSES, Arlington, VA	703 292-2310
Ms. Rebecca L. Morrison Survey Methodologist	National Science Foundation, NCSES, Arlington, VA	703 292-7794
Dr. Patricia Green Project Director	RTI International Chicago, IL	312-456-5260
Dr. Jonathan Gordon Redesign Task Leader	RTI International Atlanta, GA	770-407-4952
Mr. Peter Einaudi Data Analysis Task Leader	RTI International Research Triangle Park, NC	919-541-8765
Ms. Jennifer Pauli Data Collection Task Leader	RTI International Research Triangle Park, NC	919-485-5598
Ms. Jamie Friedman FFRDC Postdoc Survey Task Leader	RTI International Chicago, IL	312-456-5262
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. Kimberly Ault Mathematical Statistical Task Leader	RTI International Research Triangle Park, NC	919-541-7455