

Memorandum

Date: October 14, 2016

To: Margo Schwab, Desk Officer
Office of Management and Budget

From: John R. Gawalt, Director
National Center for Science and Engineering Statistics
National Science Foundation

Via: Suzanne Plimpton, Reports Clearance Officer
National Science Foundation

Subject: Request for approval of Pilot survey data collection for the Survey of Graduate Students and Postdoctorates in Science and Engineering (GSS)

The National Center for Science and Engineering Statistics (NCSES) requests approval for a Pilot study embedded within the 2016 GSS data collection cycle. NCSES is currently considering a redesign of the GSS survey to improve data utility and to potentially reduce institutional response burden.

In July and August of this year, NCSES requested and received OMB approval to conduct GSS Coordinator Survey to assess the institutional reporting practices, availability of detailed institutional data, and the awareness and use of GSS data file transfer features that could potentially reduce the burden on institutional respondents.

The GSS Coordinator Survey result will serve as the basis for the 2016 GSS Pilot survey. The Pilot survey will determine the feasibility of collecting data from the institutions through alternative data collection methods. The Pilot survey will be conducted with a stratified random sample of GSS institutions in conjunction with the regularly administered 2016 GSS. This methodological study will be conducted under the NCSES generic clearance (OMB Control Number 3145-0174).

Background

The GSS is an annual census survey of the academic institutions that offer graduate degree programs in the sciences, engineering, and health fields in the United States. Institutions are asked to report the total number of graduate students, postdoctoral appointees (postdocs), and doctorate-level nonfaculty researchers (NFRs) by demographic and other characteristic such as source of financial support.

The current GSS data collection consists of two parts. The initial data request is sent to the designated respondent (School Coordinator) at each academic institution in the fall of each academic year. In Part 1, the School Coordinator is responsible for reviewing and updating the list of all GSS-eligible organizational units for their institution (see Attachment 1, GSS Web screenshots, pp. 1-2)¹. The School Coordinator updates the list of eligible organizational units via the GSS web system. In Part 2, the School Coordinator provides the count data on the graduate

¹ Organizational units are defined as departments, programs, health-care facilities and research centers.

students, postdocs, and/or NFRs for each unit reported in Part 1. In some cases, the School Coordinator delegates Part 2 to the organizational unit respondents for completion (see Attachment 1, GSS Web screenshots, pp. 3-12).

Proposed Changes

The Pilot data collection will not ask new questions but the method used to collect the graduate enrollment and financial support data will change for a sample of institutions included in the Pilot study. The resulting data from the Pilot will be evaluated and aggregated to the same level as the regularly administered GSS for the official data release.

The changes being tested in the 2016 Pilot data collection include:

- Separate reporting of enrollment and financial support data for master's and doctoral students
- Collecting data based on the Classification of Instructional Programs (CIP) codes as a disciplinary field, instead of GSS codes
- Expanding the institutional use of file transfers for data submission instead of manual entry of data in a GSS web survey

Separate reporting of master's and doctoral student data. Currently, the GSS collects aggregated data on graduate students that cannot be parsed by degree level. Through the GSS Data User Workshop conducted in April 2016 and the Human Resource Expert Panel meeting in June 2016, NCSES received requests to collect granular data on graduate students. Specifically, the requests were to collect separate data on master's and doctoral level student enrollment and financial support data to significantly increase the usefulness and value of the GSS data.

Collecting data based on CIP codes. The collection of GSS data by CIP codes can potentially reduce response burden because these codes are commonly used at institutions. CIP is the academic field taxonomy used by the U.S. Department of Education for the Integrated Postsecondary Education Data System (IPEDS), a mandatory reporting requirement for institutions receiving Title IV funding. The GSS Coordinator Survey results indicated respondents' ability to provide GSS data using CIP codes for graduate student demographic and financial support information. Schools will have the option of using either CIP codes or GSS codes for reporting postdocs and NFRs.

Expanding the institutional use of file transfers. Currently, GSS provides two options for institutions to submit requested data. The most common method uses a series of web-forms in which institution coordinators enter data for each organizational unit. The second method allows coordinators to upload some or all of the requested data through a pre-formatted file. The latter method is of considerable utility to institutions with a large number of reporting units: rather than filling out the web tables for each reporting unit, the coordinators can prepare and upload just one aggregated file. The separate reporting of master's and doctoral students as well as the use of more detailed CIP codes will increase the number of web-forms institutions will need to complete. To mitigate this increase in burden, the Pilot survey will request that the school coordinators prepare data files and use the upload feature to submit the requested data. These transfers reduce the burden of completing what might be twice as many web forms for reporting graduate students (one for master's and another for doctoral degree students).

The feasibility of these changes was explored through the GSS Coordinator Survey to 2015 GSS institutions (see Attachment 2). A total of 840 coordinators were invited to take the survey, and 676 participated, for a total response rate of 80.5%. The survey included items on the feasibility

of separate reporting of master’s and doctoral student data and the feasibility of using CIP codes to report data by academic discipline.

Proposed Methodology

The 2016 GSS Pilot survey will be conducted in conjunction with the regularly scheduled 2016 GSS. A stratified random sample of 80 coordinators that responded to 2015 GSS will be selected for the Pilot survey. Table 1 presents information on the number of coordinators in each strata as a percentage of the eligible coordinator population.

The sample will consist of:

- 15 coordinators that used the data upload tool in the 2015 GSS
- 25 coordinators that report both master’s and doctoral students, that have a large number of reporting units (more than 15 units) and did not use the data upload tool in 2015
- 25 coordinators that report both master’s and doctoral students, that have a small number of reporting units (15 or less units) and did not use the data upload tool in 2015
- 15 coordinators that report only master’s students that did not use the data upload tool in 2015

Table 1: Sample of Pilot Institution Coordinators

Institution Coordinators	Total Number ²	Sample size	% sampled
Uploaded data in 2015	65	15	23.1
More than 15 units, master’s and doctorate, did not upload data	207	25	12.1
15 units or fewer, master’s and doctorate, did not upload data	194	25	12.9
Master’s only, did not upload data	293	15	5.1
Total	759	80	10.5

The 80 coordinators selected for the Pilot will not be part of the regular GSS data collection (see Attachment 3 for Pilot survey invitation email). The Pilot survey participants will be provided with variable and file specifications, and data aggregation macros in an Excel file that they can use to report their institution’s data at the organizational unit-level to create an upload file that will be transferred to NCSES GSS contractor, RTI, via a secure web-connection (see Attachment 4). These transfers will allow NCSES to report data by CIP codes and by GSS unit—allowing the Pilot survey data to be merged with the main 2016 GSS data. In some instances, institutions may prefer to upload a de-identified student/postdoc unit-record file. In such instances, GSS project staff will construct the appropriate data file from these submissions.

The Pilot survey data collection will include:

- Student demographic and financial support data using CIP codes instead of GSS codes as the organizing taxonomy
- Student demographic and financial support data bifurcated by degree-level to capture master’s and doctorate data separately

² Pilot sample excludes the coordinators who are only responsible for reporting postdoc and NFR data at their institution since the study is primarily focused on separate data reporting of master’s and doctoral students.

NCSES will provide training to Pilot survey participants through webinars, providing detailed instructions on preparing data files for upload. Pilot survey participants will also benefit from a set of best practices from GSS institutions that have prior experience using the data upload tool. During the Pilot survey, the RTI will maintain a dedicated help-desk to support institutions as they compile and report requested GSS data. Participants who are unable or unwilling to provide data using the pilot protocols will be allowed to revert to regular 2016 GSS collection process. NCSES will not replace these institutions in the sample. Based on the GSS Coordinator Survey response rate, we anticipate at least 80% of invited institution coordinators will participate in the Pilot data collection.

Schedule of Activities

The tentative schedule for the GSS Pilot survey is as follows:

Proposed Date	Activity/Deliverable
October 14, 2016	OMB submission for approval
October 28, 2016	OMB clearance
November 7, 2016	Send FedEx invitation to Pilot sample institutions
November 14, 2016	Send email follow-up invitation to Pilot sample institutions
November 1-18, 2016	Update GSS Instrument for Pilot survey
December 5, 2016	Launch Pilot survey
December 12-16, 2016	Pilot survey training for Coordinators
March 31, 2017	Pilot survey due date
May 31, 2017	Pilot survey Closeout and merge of Pilot data with 2016 GSS data
June 30, 2017	Draft Report
July 31, 2017	Final Report and Recommendations for 2017 GSS

Response Burden

A total of 260 hours are being requested for the 2016 GSS Pilot. Pilot institution coordinators will be invited to attend a one-hour training workshop that describes the data collection, and provides details on the data upload process. As noted above, we expect at least 80% of invited coordinators will participate in the Pilot. However, we used a 100% response rate to estimate the response burden.

Based on the results of the GSS Coordinator Survey (see Attachment 2, GSS Coordinator Survey Results: Executive Summary), we know that some coordinators think that this new data collection methodology will require the same or less burden as the regular method. Since the Pilot institutions will be submitting the same 2016 GSS data, we expect the increase in response burden to be minimal. We also expect the burden to vary across coordinators, depending on whether they have uploaded GSS data in the past, the number of units they report, and whether they confer both master's and doctoral degrees. Table 2 shows the estimated increase in burden for Pilot survey participants, by sampling strata.

Table 2. Projected increase in burden for Pilot Institutions by sample strata

Sample strata based on institution	Number of	Estimated	Estimated total
------------------------------------	-----------	-----------	-----------------

coordinators' reporting status in 2015 GSS	Pilot institution coordinators	marginal increase in burden hours	burden hours for collection
Uploaded data	15	1	15
More than 15 units, master's and doctoral, did not upload	25	4	100
15 units or fewer, master's and doctoral, did not upload	25	2	50
Master's only, did not upload	15	1	15
Burden hours for Pilot data collection			180
Burden hours for training (80 coordinators @ 1 hour)			80
Total burden hours requested			260

Sample of 15 coordinators that used the data upload tool in the 2015 GSS: The marginal burden for Pilot participation should be negligible. These coordinators are already providing data using the upload format. While there is additional data that will be collected through the separate reporting of master's and doctoral students, since these data are provided through an upload process (rather than through web-based forms), little to no additional burden is expected in terms of generating and reporting these data. Additionally, the use of CIP codes should ease the burden associated with data reporting.

Sample of 25 coordinators that reported both master's and doctoral students, have more than 15 reporting units, and did not use the data upload tool in 2015: These coordinators will encounter increased burden of learning to use the upload tool. The use of CIP codes rather than GSS codes when uploading may mitigate overall burden for this group. We estimate that Pilot participation to impose four additional hours of burden per coordinator, based on the time required to master the data upload tool and the amount of data that will be reported separately by degree level.

Sample of 25 coordinators that reported both master's and doctoral students, that have 15 or fewer reporting units, and did not use the data upload tool in 2015: These coordinators will also encounter increased burden of learning to use the upload tool. The use of CIP codes rather than GSS codes when uploading may mitigate overall burden for this group. Given the smaller number of units being reported, we estimate that Pilot participation to impose two additional hours of burden per coordinator, based on the time required to master the data upload tool and the smaller amount of data that will be reported separately by degree level.

Sample of 15 coordinators that reported only master's students and did not use the data upload tool in 2015: These coordinators will encounter a slight burden of learning to use the upload tool, but will not need to report separate degree level data since they do not enroll doctoral students. Again, the use of CIP codes rather than GSS codes should mitigate overall burden for this group. We estimate one additional hours of burden based on the time required to master the data upload tool provided through the training workshop.

Next Steps

We will use two criteria for measuring success of the Pilot survey as follows:

1) Number of schools using the upload data tool

The Pilot survey will be considered as success if about 60% of the Pilot institutions use the upload data tool to submit their data on masters and doctorate students by CIP code. Submission rates may vary by size of institution/number of units per school and other characteristics. NCSES expects a similar variation for the Pilot – about 70%-90% of schools with a large number of units will use the upload data tool but only about 10% of schools with a small number of units. For schools with a small number of units, using the GSS web survey may be easier than using the upload data tool.

2) Observe no long-term impact on response burden.

NCSES currently estimates GSS response burden to be 2.5 hours on average. We estimate that the initial additional burden of using the upload data tool will be about 1 to 4 hours (depending on institutional characteristics) with a decrease in response burden in subsequent years. Results from the GSS Coordinator Survey (July 2016) indicated that those who use the upload process were very satisfied with the upload tool, and noted time savings as one of the positive features of the upload. We also know that the large majority of coordinators who upload continue to do so year after year.

The Pilot and base 2016 GSS instruments will collect estimated response time. To determine the impact of the new data procedures on response burden, NCSES will compare the completion time reported by the Pilot institutions to their institution's previous year completion time, as well as those reported by the Pilot's base institutions.

Unless there are lower than expected Pilot response rates, higher response burden times and an inability to upload data, NCSES will modify the GSS instruments using the Pilot data collection methods. Internal NSF directorates, GSS stakeholders and data users (including GSS institutions) have provided strong rationale and agreement on the immense value of separate enrollment and financial support data on the master's and doctoral students in different fields. We anticipate that the initial increase in institutions' response burden will be one-time only and will be offset by the increased value of the data for a range of educational policy and decision making purposes.

Contact Person

Kelly Kang
Project Officer
Survey of Graduate Students and Postdoctorates in Science and Engineering
Human Resources Statistics Program
National Center for Science and Engineering Statistics
National Science Foundation
kkang@nsf.gov
703-292-7796

Attachment 1: GSS Web Survey Screens
Attachment 2: GSS Coordinator Survey Results: Executive Summary
Attachment 3: 2016 GSS Pilot Institution Invitation
Attachment 4: GSS Data File Specification Screens