

Survey of Graduate Students and Postdoctorates in Science and Engineering, Fall 2013

Kelly Kang
Project Officer
Human Resources Statistics Program
(703) 292-7796

Technical Notes

Survey Overview

Purpose. The Survey of Graduate Students and Postdoctorates in Science and Engineering (GSS)—conducted by the National Science Foundation’s National Center for Science and Engineering Statistics—is an annual survey of all academic institutions in the United States granting research-based master’s degrees or doctorates in science, engineering, or selected health (SEH) fields.¹ The GSS provides data on the number and characteristics of graduate students, postdoctoral researchers (postdocs), and doctorate-holding nonfaculty researchers (NFRs) in SEH fields. NSF uses the results of this survey to assess shifts in graduate enrollment and postdoc appointments and trends in financial support.

Data collection authority. The information is solicited under the authority of the National Science Foundation Act of 1950, as amended. The Office of Management and Budget (OMB) control number is 3145-0062, and expires on 30 September 2014.

Survey contractor. RTI International.

Survey sponsors. The GSS is sponsored by the National Science Foundation and the National Institutes of Health (NIH).

Key Survey Information

Frequency. Annual.

Initial Survey Year. 1966.

Reference period. FY 2013.

Response unit. Organizational units in academic institutions (e.g., schools², departments, degree granting programs, university-affiliated research centers and health care facilities).

Sample or census. Census.

Population size. 564 academic institutions.

Sample size. Not applicable.

Survey Design

Target population. The survey target population included all academic institutions in the United States and its territories (Guam and Puerto Rico) that granted research-based master’s degrees or

doctorates, appointed postdocs, or employed NFRs in SEH fields as of fall 2013. For graduate students, *field* refers to the field of the reporting unit in which the student is enrolled. For postdocs, *field* refers to the field of the unit that reports postdocs to the GSS. The survey was revised in the 2010 survey cycle to collect postdoc data and graduate student data at comparable levels of detail.

In 2013, the survey universe consisted of 564 institutions. Data were collected at the organizational unit level (e.g., departments, degree-granting programs, research centers, health facilities) and included demographic and funding information for graduate students and postdocs. Detailed information on the institutions, schools, and units is provided in tables A-3, A-4, and A-5.

Sample frame. Eligible academic institutions are identified primarily through the Integrated Postsecondary Education Data Systems (IPEDS).

Sample design. Not applicable; all eligible units are surveyed.

Data Collection and Processing Methods

Data collection. GSS data are collected through coordinators at eligible schools. Coordinators are individuals, assigned to their role by their institution, who are knowledgeable about their postdocs and graduate students. They are responsible for identifying all GSS eligible units, collecting GSS data, and submitting the data to the survey contractor. Some schools chose to assign a separate graduate student coordinator and postdoc coordinator, whereas others chose to have one coordinator report all data.

Data collection for the GSS is done in two parts. The first part identified the organizational units within the school. The second part collected counts and selected characteristics of graduate students, postdocs, and NFRs.

The survey was launched in October 2013 and concluded in May 2014. The deadline for Part 1, the update of the unit list, was 30 November 2013. Schools that missed the Part 1 deadline received special attention from the survey contractor early in the survey cycle. The deadline for submitting data for Part 2 was 28 February 2014.

Mode. In 2013, the Web survey was the primary mode of data collection. A paper worksheet was provided for informational purposes and to assist in preparing figures to be entered later in Part 2 of the Web survey. The content and format of the paper worksheet were identical to Part 2 of the Web survey. A small number of coordinators did not use the Web survey but chose to submit their Part 2 data in an Excel file. These data were loaded in the Web survey by the survey contractor. The responses were provided as follows:

- *Web.* 500 institutions (88.7%) reported using the Web survey.
- *Upload tool.* A total of 59 institutions (10.5%) uploaded some portion of the data via the upload feature.
- *Excel spreadsheet.* Five institutions (0.9%) reported via Excel spreadsheets.

Response rates. Response rates are calculated based on responses to the survey's various data-collection grids (graduate student and postdoc counts, by ethnicity and race; full-time graduate

student and postdoc counts, by primary source or mechanism of support; counts of postdocs, by type of doctoral degree and primary mechanism of support; counts of postdocs, by type of doctoral degree and citizenship; counts of postdocs, by origin of doctoral degree; and counts of NFRs, by type of doctoral degree and sex).

The method for calculating response rates for units has changed over time. From 2007 to the present, complete row and column totals for all grids and all details summing to the totals were complete responses; some data reported (e.g., only grand totals) but data incomplete for any of the grids were partial responses; no data reported in any grid was a nonresponse. For information about the methods used prior to 2007, please see the Technical Notes section of *Graduate Students and Postdoctorates in Science and Engineering: Fall 2007* at <http://www.nsf.gov/statistics/nsf10307/>. Response-rate calculations for 2007 and beyond adhere to the American Association for Public Opinion Research standards for computing response rates.³

- *Unit response.* In 2013, the GSS received complete responses from 12,056 (86%) of the 14,019 eligible units. An additional 1,923 units (13.7%) were partial respondents. The remaining 40 units (0.3%) were nonrespondents. Table A-6 shows the unit response rates from 1975 through 2013.
- *School responses.* Of the 680 schools eligible for the 2013 GSS, 675 schools (99.3%) were complete respondents (i.e., 90% or more of the school's units provided complete or partial data), 3 schools (0.4%) were partial respondents (i.e., at least 50% but less than 90% of the school's units provided complete or partial data), and 2 schools (0.3%) were nonrespondents (less than 50% of the schools units provided data).
- *Institutional response.* Institutional response rates were calculated using the same criteria for schools. Of the 564 eligible institutions, 561 institutions (99.5%) were complete respondents, 1 institution (0.2%) was a partial respondent, and 2 institutions (0.4%) were nonrespondents.

Zeros versus nonresponse. As in previous years, data collection grids in the Web survey were prefilled with zeros. Respondents were asked to mark a checkbox if the unit does not have any eligible data to report. Grids with a marked checkbox contributed to a complete response for the unit. Grids with unchanged, prefilled zeros and an unmarked checkbox disqualified the unit from complete response status.

Data editing. Data quality is ensured by interactive edit checks built in the Web survey and by a comprehensive review after the data are submitted by the coordinator. The Web survey edits verify that the data entered are internally consistent and are within an expected range, often based on the previous year's data. During follow-up, unit respondents are asked to explain the discrepancy whenever counts differ substantially from those of the previous year.

All data submitted by the academic institutions are reviewed to ensure that all data fields are complete and that data are internally consistent. These quality checks were conducted when counts remained identical to the previous year and also when the school's unit list, total counts, and distribution of counts had notable changes. Changes to the unit list included all unit additions and deletions and also changes to the highest-degree-granted status, GSS code, and unit name. All units that had total counts or a distribution of counts within a given data item that were substantially

different from the previous survey cycle were reviewed. Any units with cell counts that are more than 20% above or 20% below its corresponding prior-year data are flagged for edit verification.

Data fluctuations that were not sufficiently explained by the comments provided by the respondents during data collection were flagged for follow-up by e-mail or telephone call to the coordinator. Revisions were made directly in the Web survey by the coordinator, unit respondents, or GSS contractor staff at the direction of the coordinator. See "Survey Quality Measures" below for a discussion of the types of measurement error detected in the 2013 data review and follow-up process.

Imputation. In the GSS, all missing data are subject to imputation. For units missing a key total (e.g., total part-time students, total postdocs, or total NFRs) with at least 1 year of qualified historical data, a carry forward imputation method was used. Inflation factors were calculated for the key totals to account for year-to-year change. The previous year's key totals were then multiplied by these inflation factors to calculate the imputed values for the current year's key totals.

All other variables were imputed by distributing the imputed key totals according to the previous year's proportions. For units that reported totals but no details, details were imputed according to the prior distribution if qualified historical details were available. Otherwise, nearest neighbor imputation techniques were used. In this method, a donor unit that was "nearest" to the unit whose data were being imputed (imputee) was identified among all responding units having similar characteristics as the imputee (such as having the same GSS code for program fields and offering a doctoral degree). When graduate student details were imputed, the nearest neighbor selected had full-time and part-time graduate enrollments that were most similar to the imputee's enrollments. The imputed values were calculated by adjusting the donor's values to account for the difference in full-time and part-time enrollment totals between the two units.

Similarly, when postdoc or doctorate-holding NFR details were imputed, the total number of postdocs or NFRs, respectively, was used to choose the nearest neighbor. If the postdoc or NFR total was missing, the graduate student totals were used to select the nearest neighbor to impute the postdoc or NFR variables. If either the postdoc or NFR key total (or both) was missing, other available key totals were used to select the nearest neighbor to impute the data. The same donor was then used to impute the details corresponding to the imputed key totals.

In rare circumstances when no graduate student data were available from a new unit, IPEDS completions and enrollment data were used to estimate graduate student totals. Based on the imputed totals, the details were then imputed by the nearest neighbor method mentioned above. Because IPEDS does not collect data on postdocs and doctorate-holding NFRs, a nearest neighbor was selected from the 2013 GSS data to estimate these counts, if necessary, using the graduate student totals to select a donor. For units in institutions that had not been in the GSS before, postdoc and NFR values were imputed as zero rather than using IPEDS-based imputation.

Tables A-7 to A-15 show the counts for imputed data and imputation rates.

Missing values from the postdoc and NFR data collection that began in 2010 were imputed for the first time in 2012 using detailed, research-based imputation procedures.⁴ The 2010 and 2011 postdoc data updated in 2012 supersede all previously released data.

Weighting. Not applicable.

Variance estimation. Not applicable.

Survey Quality Measures

Sampling error. Not applicable.

Coverage error. The universe of higher education institutions is regularly reviewed to identify new, potentially eligible institutions to minimize the coverage error.

Nonresponse error. The GSS is subject to nonresponse error. Data review and follow-up indicated that zeros reported by respondents sometimes represent nonresponse rather than actual zero counts. Not distinguishing the two could result in low estimates, given that data for a given variable are not imputed when item nonresponse is misinterpreted as a zero response. Prior to 2007, prefilled zeros were considered legitimate responses if the grid was left with all zeros in place. Beginning with 2007, true zeros reported by the respondents were distinguished from those remaining from nonresponse by a checkbox indicator, which was added to explicitly confirm zeros for the grid prior to submission. In 2010, the first-time, full-time graduate student cells were the only cells with the potential for ambiguous zero counts; the remaining cells either had checkboxes to confirm a zero entry or were not prefilled with zeros. In 2011, a checkbox was added for first-time, full-time graduate students to eliminate ambiguous zero counts.

Measurement error. The GSS is also subject to measurement error that arises when variables of interest cannot be precisely measured. Review of the data, cognitive interviews, usability tests, pilot tests, site visits and other methodological activities with the institutions have pointed to a number of possible sources of measurement error. Possible sources of measurement error are listed below.

- *Double counting.* Anecdotal evidence indicated that some misreporting may have occurred when an institution had more than one coordinator or offered joint programs. To reduce double counting, facilitate communication, and allow sharing of reported data, a screen in the Web survey provides names and contact information for all school coordinators at the institution.
- *Inclusion of practitioner degrees.* Graduate students working toward practitioner degrees, particularly in health fields with explicit exclusions may sometimes be overreported. Starting with the 2007 survey cycle, survey materials indicated that students pursuing master's, DDS, MD, and certain other degrees in specified fields should be excluded from the counts. During the imputation process, new units that were suspected of having reported graduate students in excluded degree-field programs based on the GSS code were set to zero graduate students to be conservative, in the absence of other information. In the 2011 survey cycle, checks were built into the Web survey to remind respondents to exclude students pursuing practitioner-based degrees.
- *Difficulty in reporting source and mechanism of support.* Feedback from respondents and methodological research indicated that financial support data were oftentimes difficult for respondents to report. Difficulties in reporting these data may occur because the information may not be stored in one centralized database; financial support may not always be channeled through the institution (e.g., self-support); and foreign sources of support may not

always be known. Respondents may also have difficulty categorizing financial information by field, such as when a student is enrolled in one unit but receives support from another. Therefore, these data may be more prone to measurement error than other survey data items. Finally, institutions define mechanisms of support differently (e.g., fellowships vs. traineeships) and may report individuals according to the institution's definition rather than that provided by the GSS. The grids now include “unknown” categories, beginning with the 2010 survey.

- *Difficulty in reporting postdocs and NFRs.* Many respondents indicate in the Web survey that they are unable to provide data on their unit's postdocs or NFRs because they do not know all of the units that employ postdocs and NFRs. Starting with the 2010 survey cycle, schools were given the option of appointing a separate postdoc coordinator who may be more knowledgeable about the postdocs or NFRs at his or her school to provide these data

Data Comparability (Changes)

Changes in survey coverage and population. No revisions affected the survey universe in 2013; revisions to the eligibility criteria for units and fields of study in prior survey cycles are described below.

- *Fields of study.*

2011: The GSS code list was updated from the 2000 Classification of Instructional Programs (CIP) taxonomy to the 2010 CIP taxonomy. Some new CIP fields were added to the GSS codes, some CIP fields were moved between GSS codes, and some CIP fields were removed as ineligible fields. The impact on field-level counts was typically small and did not change the overall trend from 2010 to 2011. A crosswalk between the 2010 CIP codes and the GSS codes are provided in table A-16. For more details on the taxonomy changes in 2011, see Technical Notes at <http://www.nsf.gov/statistics/nsf13331/>.

2007: GSS-eligible, degree-granting programs were updated from the 1990 to the 2000 CIP taxonomy. Degree-granting programs were specified by a six-digit CIP code rather than a four-digit CIP code. Three newly eligible fields were added, some degree-granting programs became ineligible, and others were reclassified. Practitioner-based fields were deemed ineligible. For more details, see Technical Notes at <http://www.nsf.gov/statistics/nsf10307/>.

- *Units.* Institution follow-up procedures were introduced in 2007 to improve the institution reporting of eligible units and the exclusion of ineligible units. In the 2013 GSS, changes in the net number of units increased by 67 (see table A-1). For more details on changes introduced in 2007, see Technical Notes at <http://www.nsf.gov/statistics/nsf10307/>.
- *Eligibility and degree-granting status.* Institutions are classified as doctorate granting if at least one GSS-eligible unit confers doctoral degrees. In 2013, five institutions changed GSS degree-granting status, and two institutions changed their institution name. The status of one institution changed from eligible to ineligible based on criteria for inclusion in the GSS (see table A-2).

Changes in survey content.

- *Sex.*

2010: Began collecting citizenship, ethnicity, and race data on postdocs by sex and type of doctoral degree data on NFRs by sex.

2008: Began collecting the number of first-time, full-time male graduate students by ethnicity and race; full-time male graduate students by source of support; male postdocs by source of support; and male doctorate-holding NFRs. Previously, the number of men was inferred by subtracting the number of women from the total.

- *Ethnicity and race.*

2010: Began collecting ethnicity and race data for postdocs who are U.S. citizens and permanent residents using the same categories as used for graduate students.

2008: Revised ethnicity and race categories to correspond to IPEDS by combining "Hispanic/Latino, One race only" and "Hispanic/Latino, More than one race" categories into "Hispanic/Latino (one or more races)."

- *Citizenship.*

2010: Began collecting citizenship data on postdocs using the same categories as used for graduate students. In previous years, only counts of postdocs who are foreign nationals holding temporary visas were collected.

2008: Clarification made for "non-U.S. citizens" to exclude non-U.S. citizens residing outside of the United States who are enrolled in an online degree program at a U.S. institution.

- *Financial support.*

2010: Began collecting data on the largest source of financial support and on the largest mechanism of support separately for postdocs. For mechanism of support, "nonfederal sources" was replaced with "other support."

2008: Graduate student data no longer collected for NIH teaching assistantships because NIH does not offer financial support for students through this mechanism.

2008: Began collecting number of full-time graduate students whose largest source of support came from a non-U.S. source via teaching assistantship.

- *Doctoral degree.*

2010: Began collecting more detailed information on postdocs' and NFRs' doctoral degree type. Categories were added for those holding a doctoral degree (e.g., PhD, ScD, DEng), a professional degree (e.g., MD, DVM, DO, DDS), and dual degrees (MD-PhD, DVM-PhD), as well as for those for whom type of degree was unknown. In the previous years, the

degree-type question simply asked the respondents to report the postdocs with MD, DO, DDS, DVM from the total postdoc count.

2010: Began collecting postdocs' doctoral degree type by citizenship and by country of origin (United States, foreign, unknown) of doctoral degrees. Also began collecting NFRs' doctoral degree type by sex.

Changes in Web survey instrument. Modifications were made to two sections of the 2013 GSS Web survey instrument as follows:

- *Unit respondent update.* In prior rounds, coordinators occasionally had chosen the option to edit a unit respondent's name when they actually intended to replace the unit respondent with a different contact; sometimes, this resulted in confusion about an individual's role and login information. Therefore, in the 2013 GSS, a dialog box was added that reminded coordinators to use the "Change Respondent" feature if their goal was to assign a different contact. The "Change Respondent" feature allowed coordinators to select from a pool of existing contacts or add a new contact person.
- *File upload feature.* Prior to 2013, it was required that all rows for a question be included in the upload file for a given unit, regardless of whether the line contained all zero data. Starting in 2013, it was no longer required for the user to provide zero lines, as long as the detail information added up to the total line. Also, coordinators who chose to use the upload feature could download a simple unit listing that contained the GSS unit identification, unit name, GSS code, and highest degree granted and use that to build their crosswalk to their own systems. Finally, the upload instructions were revised so that the terminology and language were easier to follow for respondents.

Changes in survey procedures.

- 2013: Three different versions of the launch e-mail were used. One version was sent to coordinators who used the data file upload feature in 2012, informing them of updates. An alternate version was sent to coordinators who might benefit from the upload (i.e., coordinators with a large number of units and who were not currently using the upload feature). The third version was the standard launch e-mail, with no mention of the upload. In addition to the launch e-mails, the data review and retrieval efforts began in January, earlier than in prior years.
- 2010: Significant effort was made to ensure that appropriate personnel were providing postdoc and NFR data. As a result, it is unclear how much of the increase reported in 2010 represented actual growth in postdocs and how much resulted from improved data collection. More information on the improved data collection and changes in postdoc data is available at <http://www.nsf.gov/statistics/gradpostdoc/>.

Historical changes. Changes have been made over the years to the coverage and content of the GSS to keep it relevant to the needs of data users. Such changes prevent precise maintenance of trend data. For more information on the past-year changes, see Technical Notes at <http://www.nsf.gov/statistics/nsf10307/>.

Definitions

Ethnicity and race. The GSS uses definitions of ethnicity and race that are based on the OMB's Standards for the Classification of Federal Data on Race and Ethnicity:

- *American Indian or Alaska Native.* A person of only one race having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment.
- *Asian.* A person of only one race having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent—for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
- *Black or African American.* A person of only one race having origins in any of the black racial groups of Africa.
- *Hispanic or Latino (more than one race).*⁵ All individuals of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. This category includes individuals who are Hispanic or Latino and any other race.
- *Not Hispanic or Latino.* Individuals who are not of Hispanic or Latino descent, regardless of race.
- *Native Hawaiian or Other Pacific Islander.* A person of only one race having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific islands.
- *White.* A person of only one race having origins in any of the original peoples of Europe, the Middle East, or North Africa.
- *More than one race.* A person of two or more of the race categories listed above.
- *Unknown ethnicity or race.* A person whose ethnicity or race is unknown or not stated.

First-time. Students enrolled for credit in a graduate degree program in an organizational unit for the first time in fall 2013. This may include graduate students previously enrolled in another graduate degree program at the institution or at another institution. It may also include students who already hold another graduate or professional degree.

Full-time and part-time. Respondents were instructed to use their institution's definition.

Graduate student mechanisms of financial support.

- *Fellowship.* A competitive award (often from a national competition) given to a graduate student that requires no work of the recipient.
- *Research assistantship.* A financial award given to a graduate student where most of the student's responsibilities are devoted primarily to research.

- *Teaching assistantship*. A financial award given to a graduate student where most of the student's responsibilities are devoted primarily to teaching assistant activities.
- *Traineeship*. A financial award given to a graduate student selected by the institution.
- *Other support*. All other mechanisms of support for graduate students.

Graduate student source of financial support.

- *Federal sources*. Financial support provided by the federal agencies. Excludes federally guaranteed student loans.
- *Nonfederal sources*. Financial support from state and local government; institutional support, such as tuition waivers and stipends; support from foreign sources, such as foreign governments, foreign firms, and agencies of the United Nations; and support from other U.S. sources, such as nonprofit institutions, private industry, and all other nonfederal U.S. sources.
- *Self-support*. Supported by loans (including federal loans) or personal or family financial contributions.

Historically black colleges and universities (HBCUs). Institutions of higher education that have been historically considered to enroll predominantly black students. The Department of Education maintains an official list of HBCUs, which is reviewed annually.

Mechanisms of financial support for postdocs.

- *Traineeship*. A financial award given to a postdoc selected by the institution.
- *Research grant*. A financial assistance award given to an organization or an individual postdoc that supports specific research goals.
- *Other support*. All other mechanisms of support for postdocs.

Nonfaculty researchers (NFRs). All doctorate-holding researchers who (1) are not considered either postdoctoral researchers or members of the faculty and (2) are involved principally in SEH research activities. Also referred to as “other doctorate-holding nonfaculty researchers.”

Postdoctoral researchers (postdocs). The definition of a postdoc varies by institution. Respondents were instructed to use their institution's definition of a postdoc. NSF defines a postdoc as meeting both of the following qualifications: (1) holds a recent doctoral degree, generally awarded within the last 5–7 years, such as PhD or equivalent (e.g., ScD, DEng), or first-professional degree in a medical or related field (e.g., MD, DDS, DO, DVM), or foreign degree equivalent to a U.S. doctoral degree and (2) has a limited-term appointment, generally no more than 5–7 years, primarily for training in research or scholarship and working under the supervision of a senior scholar in a unit affiliated with the institution.

Sources of financial support for postdocs.

- *Federal sources.* Financial support provided by the federal agencies.
- *Nonfederal sources.* Financial support from state and local government; institutional support; support from foreign sources, such as foreign government, foreign firms, and agencies of the United Nations; and other U.S. sources, such as support from nonprofit institutions, private industry, and all other nonfederal U.S. sources.
- *Personal resources.* The personal and family financial resources, including federal and other loans.
- *Unknown or not stated.* Sources of financial support for the postdoc are unknown or cannot be determined.

¹ The research doctorate is a research degree that (1) requires an original contribution of knowledge to a field (typically, but not always, in the form of a written dissertation) and (2) is not primarily intended for the practice of a profession. For additional survey information and available data related to graduate student enrollment and postdocs in science and engineering, see <http://www.nsf.gov/statistics/srvygradpostdoc/>.

² In this report, the term *school* refers to a graduate school, medical school, dental school, nursing school, or school of public health; an affiliated research center; a branch campus; or any other organizational component within an academic institution that grants an SEH degree.

³ See response rate 3 calculation in American Association for Political Opinion Research (AAPOR). 2011. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 7th ed. Page 45. Deerfield, IL: AAPOR.

⁴ See Harter R, Wang J, Einaudi P, Green P. 2013. GSS Imputation Research. Report prepared for the National Science Foundation. Research Triangle Park, NC: RTI International.

⁵ The OMB standards designate Hispanics as an ethnic group rather than a racial group. Following these standards, Hispanic is not counted as a race in GSS. Cognitive interviews with respondents have shown that this is a source of considerable confusion. For example, black Hispanics and white Hispanics may be counted as "Hispanic, More than one race" rather than "Only one race, Hispanic." The ethnicity and race categories were aligned to IPEDS by combining the "Hispanic/Latino, More than one race," and "Hispanic/Latino, One race only," categories. In 2008, these two Hispanic categories were collapsed into one: "Hispanic/Latino ethnicity (one or more races).