

# **SF-83 SUPPORTING STATEMENT**

**for**

**Survey of Earned Doctorates**

**2018 and 2019 survey cycles**

**Section B**

## SECTION B: Collection of Information Employing Statistical Methods

### B.1. Universe and Sampling Procedures

The SED is a census of all students receiving a research doctorate between July 1 and June 30 of the following year. All institutions identified in IPEDS as granting doctoral degrees are asked to participate *if*: (1) they confer “research doctorates” and (2) they are accredited by one of the regional accreditation organizations recognized by the U.S. Department of Education. The participating schools distribute the link to the online questionnaire or distribute the paper questionnaires to their research doctoral recipients at the time of graduation. The SED maintains the universe of research doctorate-granting institutions each year by comparing the list of institutions from IPEDS against the schools participating in the SED. If a new institution is found to be offering a research doctorate, the institution is contacted and, based on eligibility criteria, added to the SED universe.

<b>Academic Year (1 July – 30 June)</b>	<b>Number of Institutions in Universe</b>	<b>Number of Doctoral Graduates</b>	<b>Response Rate*</b>
2014	433	54,070	90.6%
2015	445	55,006	90.2%
2016	445	~56,000	~90%
2017	445	~57,000	~90%
2018	~450	~58,000	~90%

\*This response rate represents the rate at which doctorate recipients complete and return SED questionnaires.

A high response rate is essential for the SED to serve its role as the frame for the Survey of Doctorate Recipients (SDR), and as the only reliable source of information on very small populations (racial/ethnic minorities, women, and persons with disabilities) in specialized fields of study at the doctoral level.

The feasibility of conducting the SED on a sample basis, and the utility of the resulting data, have been considered and found to be inadequate. Many institutions participate in the survey to receive comprehensive information about all of their research doctorate recipients and to make comparisons with peer institutions. Experience indicates that the roughly 590 ICs that voluntarily distribute the SED would have great difficulty carrying out a sampling scheme. The current process is easy for ICs given that schools often refer their students to an online graduation checklist, where the SED is but one step in the graduation process. In addition, conducting the SED on a sample basis would produce poor estimates of small populations (e.g., racial/ethnic minorities) earning degrees in particular fields of study, and such data are important to a wide range of SED data users.

A second sampling option – a mailing to doctorate recipients *after* graduation – would likely result in much lower response rates. Obtaining accurate addresses of doctorate recipients is very difficult, particularly for the foreign citizens who represent an ever-growing proportion of the

doctorates recipient universe each year. Although universities could help, they would incur the additional burden of determining current addresses for doctorates who have graduated. This is a somewhat ineffective process because the addresses of new doctorates are outdated almost immediately after graduation.

A third alternative, sending the questionnaire to doctorate recipients at a selected subset of institutions, would result in only a marginal decrease in respondent burden because the largest universities, all of which would need to be included in such a scheme, grant a disproportionate number of doctoral degrees. For example, the 50 largest institutions of the 450 total annually grant slightly over 50 percent of all doctoral degrees. Application of these sampling techniques would reduce both the utility of the data and the overall accuracy of the collected data.

Sampling doctorates would decrease data quality while increasing burden on the graduate schools that administer the survey, thereby decreasing the incentive for institutions to participate.

Given that the SED is a census, weighting is not conducted. Missing information about non-responding individuals is obtained, where possible, from public records, commencement lists, etc. Both unit and item nonresponse are indicated by including categories of “unknown” for all variables in tabulated results.

## ***B.2. Survey Methodology***

The SED questionnaire is distributed to new research doctorate recipients by approximately 590 independent programs within approximately 450 doctorate-granting institutions in the United States. The SED (either web or paper) is filled out at the time the individuals complete all requirements for their doctoral degrees. If paper questionnaires are completed, ICs return them to NSF/NCSES’s survey contractor. Because doctorate recipients complete the requirements for graduation throughout the year, the questionnaire distribution and completion process is continuous.

The institution (usually the graduate dean’s office) is the main SED interface with the doctorate recipient and experience shows that the interface is highly effective. The distribution of the questionnaire by the university itself, the clear nature of the questionnaire, and the cooperation of the graduate deans all combine to keep survey response rates at 90 percent.

When the completed paper questionnaires are received by the survey contractor, they are edited for completeness and consistency and then entered directly into a computer-assisted data entry (CADE) program. Surveys received via the web survey mode do not need to be data entered and are edited mainly through a series of pre-programmed skip patterns and range checks, which allow many errors to be corrected immediately.

The survey contractor sends an Address Roster to ICs to obtain contact information for nonresponding students. The survey contractor also uses web-based locating sites to identify contact information for nonrespondents. A series of emails or postal letters is sent to graduates

who did not complete the survey through their graduate school. They are requested to participate and sent a PIN/password for web access (see Attachment 9 for a sample letter).

Finally, nonrespondents are given the opportunity to complete a shortened version of the survey over the telephone. If, by survey close-out, an individual has not responded, public information from commencement programs or other publicly accessible sources is used to construct a skeletal record on that individual. The institution may also be asked to help provide data to complete skeletal records for these nonrespondents. The skeletal record contains the name, PhD institution, PhD field, degree type, calendar year that the doctorate was earned, month that the doctorate was earned, and (usually) the sex of the doctorate. If a survey questionnaire is later received from a previous nonrespondent, the skeletal record is replaced by the information provided by the respondent.

### ***B.3. Methods to Maximize Response***

The SED has traditionally obtained a high response rate, with an average of above 90% over the past 30 years. It owes this high rate, in part, to the use of the data by the ICs/graduate deans, who go to great lengths to encourage participation on the part of their graduates. Soon after the data are released each year, each graduate dean receives a profile of their graduates, compared with other institutions in their Carnegie class.

In addition to the importance the universities themselves place on the data, the high response rate is also due to university outreach efforts pursued by NSF/NCSES. Throughout the data collection period, school participation is constantly monitored. Doctorates awarded each commencement date are compared to data from the previous round, and fluctuations in expected returns are flagged. Schools with late returns or reduced completion rates are individually contacted. Staff site visits, primarily to institutions with low response rates, are also critical to maintaining the SED's consistently high response rate.

Along with the broad efforts to maintain high response, targeted efforts to prompt for missing surveys and critical items are also key. The survey contractor works with ICs and also uses web-based locating sites to contact students by email and mail. A series of contacts is sent to any graduate who did not complete the survey through their graduate school, requesting their participation and including a PIN/password for web access, as well as a paper questionnaire when postal mailings are sent. Additionally, nonrespondents are given the opportunity to complete a slightly shortened version of the survey over the phone.

Finally, a Missing Information Roster (MIR) is sent to ICs who can sometimes provide critical item information (sex, race/ethnicity, citizenship, etc.) in addition to addresses. The results of these varied efforts significantly increase the number of completions as well as reduce the number of missing critical items, thereby improving the quality of SED data.

Institutional and individual response rates are evaluated annually. Institutions with poor response rates are targeted for conference calls or site visits to discuss their procedures and potential

improvements for achieving a higher response rate. Such efforts typically have been successful in raising response rates.

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

In recent years, NSF/NCSES has conducted extensive review of the SED survey methods and testing of the SED questionnaire. Below is a list of the major activities that have been conducted since the previous OMB clearance submission (see Attachment 10 for a list of recent methodological studies).

#### ***Data Collection and Processing Tests***

During data collection, item non-response is continuously monitored to ensure that emerging problems can be identified early and appropriate corrective measures implemented. Completed questionnaires are constantly compared with the universities' graduation lists and commencement programs to ensure that only questionnaires for persons with earned research doctorates are included. Quality control checks related to processing for paper and electronic questionnaires and missing information are also continuous. Overall, NSF/NCSES conducts a continuous evaluation of the accuracy of SED coding, editing, and data entry processes. The results consistently indicate that the error rate is very low (less than one percent).

#### ***Experiments***

Strategies to prompt survey completion are continually examined to maximize response rates and reduce survey costs. During the 2015 data collection cycle, two experiments were conducted to test improving response rates. First, the inclusion of a progress bar in the web survey was tested to determine if it would reduce the number of breakoffs and, ultimately, lead to more completed surveys. Nonrespondents were contacted through follow-up efforts and selected for the experiment. Treatment group members saw a progress bar that displayed their advancement through the survey both by visual increase of the bar and by indicating the percentage of the survey completed. Control group members did not see a progress bar, in keeping with the past SED web survey design. The results indicated that the progress bar had a small positive effect on single session and final completion rates, but the effect was not always statistically significant. These inconsistent results may have been due to the 25-minute (average) length of completing the 2015 SED web survey. This 25-minute length is midway between the length (20 minutes) at which research has found progress indicators reduce break-offs and the length (30 minutes) at which progress indicators have been found to increase break-offs. The redesigned 2017 SED web survey takes 20 minutes (on average) to complete and includes a progress indicator. Break-offs and time-to-complete by page are currently being examined.

The second experiment tested contact strategies for nonrespondents. This experiment tested the effectiveness of starting with email prompts rather than mail prompts for SED nonrespondents for whom the doctorate-granting institutions had provided both a mailing and an email address. Past follow-up protocol for all SED nonrespondents with a mailing address was to send them up to five mail prompts (four letters and one postcard) before any other contact was tried, regardless

of the presence of an email address in the sample database. After the five mail contacts, nonrespondents were sent up to two email prompts. This strategy was based on the idea that mailing addresses supplied by institutions should be used as quickly as possible before graduates relocated, whereas email addresses would still be effective after graduates relocated. For the experiment, nonrespondents selected for the treatment group were first sent a series of two emails followed by the series of five mail prompts.

The results demonstrated that regardless of whether emails or mail prompts were sent first, eventually the same response rate could be achieved. Email before mail resulted in cost benefits, as doctorate recipients receiving emails were more likely to complete the survey sooner, thus requiring fewer contacts. Sending fewer contacts reduces operational costs (both in labor and materials, given that emails are less costly to send than postal mailings). Additionally, receiving their responses sooner allows for more data processing time, potentially resulting in higher quality data.

Though the experiment did not measure respondent burden, it is likely that email prompts have the additional benefit of reducing burden by making it easier for respondents to access the survey. Sending a paper invitation for a web survey requires that the letter include a URL address for the survey, which recipients must manually type into a computer (as opposed to clicking on a link in an email). Considering that the web is the dominant SED completion mode (2015 SED 93%), when it is possible to obtain email addresses, email prompts will be sent prior to mail prompts.

### ***Survey Quality Tests and Research***

Several tasks were completed since the last OMB package, including tasks that informed the recommendations for the next cycle. These tasks ranged from continuous assessments of everyday processes to overarching reviews of the institutions and degrees included in the survey to confirm the completeness and accuracy of the SED universe.

The following tasks are conducted regularly throughout each survey cycle:

- Review of systems, programming, data preparation, and quality control processes with a goal of shortening data collection and delivering the final data file earlier.
- Merging data on a flow basis to identify and correct data inconsistencies and to reduce the amount of time between the close of data collection and the data release.

These tasks are completed annually, prior to the beginning of data collection or the start of data preparation:

- Comparison of the IPEDS database of doctorate-granting institutions to the SED universe to identify institutions newly offering doctorate programs that are not currently in the SED. Based on this review six new institutions were deemed eligible for participation in the 2015 SED and eight for the 2016 SED.
- Review of the IPEDS database and SED results to determine if any institutions that currently participate in the SED are offering eligible degrees that the SED has not yet acknowledged. Based on this review, six programs at institutions already in the SED

were deemed eligible for participation in the SED for the 2015 round and five programs for the 2016 round.

- Discussion of possible improvements in the coding and editing processes to ensure faster data entry resulting in timelier follow-up with nonrespondents.
- Consultation with data processing managers on issues of paper and electronic data handling and mergers.
- Analysis of confidentiality issues, particularly of data products that will be publicly available.
- Coordination of items common to the SDR instrument (see section A.4).

The following tasks are completed annually at the end of each data collection period. The results are compiled and reviewed before each new OMB clearance cycle to inform possible changes:

- Extensive reviews of unit and item-by-item frequencies and item analysis for floor and ceiling effects (i.e., whether quantitative response options go low enough and high enough for the range of SED responses).
- Review of all respondent comments for concerns over confidentiality or item improvements.
- Review of “other, please specify” information in consideration of expanding or changing response options.

The following tasks are typically conducted biennially:

- Detailed review of emerging and declining fields of study. The result of the review completed in preparation for the 2018 SED is the addition of three new fields to the SED field of study taxonomy, the renaming of two fields, and the removal of one field.
- Review of any new non-PhD doctorate degree to confirm that it is a research degree and thus eligible for inclusion in the SED. Based on this review in 2014, it was determined that a newly offered Doctorate of Design (DDes) was eligible for the SED.

Finally, the following specialized studies were conducted during the last OMB cycle:

- **Redesign of the SED Data Collection Instrument:** As part of the ongoing effort to reduce respondent burden, increase efficiency, and improve the quality and utility of SED data, NCSES conducted a research effort to redesign the data collection instrument. Based on past SED research, the survey research literature, and data from previous cycles, revisions were made to all sections of the instrument. Special attention was paid to perceived redundancies in item content, terms that seemed to be poorly understood, and inefficiencies and reporting errors related to item flow. Most notably, items gathering educational history information of the doctoral recipients were restructured to align with respondents’ recall of the start and conferral dates of each of their postsecondary degrees. This clarification is important as the SED relies on these dates to compute key time-to-degree measures. In addition to restructuring data items, a few new questions were added to address new areas of interest and expand the analytic utility of the data, and a few items were retired. Multiple rounds of cognitive testing were conducted with doctoral graduates to refine the items and flow. The final recommendations that were non-substantive were incorporated in the 2017 SED instrument. The substantive changes (i.e., addition of items and removal of items) are scheduled to be incorporated in the 2018 SED.

- **Defining and Measuring Interdisciplinary Research:** This study examined the value and feasibility of collecting information about interdisciplinary doctorate degree (IDD) programs and interdisciplinary dissertation research (IDR). A research team reviewed relevant literature to gain a basic understanding of how IDD and IDR are defined and interrelated, examined published sources on IDD currently supported by U.S. universities, synthesized the published typologies of IDDs, and analyzed the DRF data on dissertation fields of study to identify institution types, particular institutions, and fields of study that might be supporting IDR. NSF/NCSES is considering follow-up research that would examine alternative ways to identify IDD programs at particular institutions, and would analyze the validity and scalability of these alternatives.
- **SED Institution Exit Interview Review:** This study examined the use of exit surveys by a sample of 50 SED institutions. The study cataloged the range of content included in institutional exit surveys and provided an initial estimate of the proportion of all SED institutions that currently administer exit surveys to their graduating doctoral students. As a follow-up to this study, a research plan has been developed to collect information from all SED institutions about how data on graduate school experiences are currently used and whether the inclusion of such items on the SED would be valuable to them (see the next section, *Proposed Tests and Research*, for further details).

### ***Proposed Tests and Research***

Over the course of the upcoming OMB clearance cycle, NSF/NCSES anticipates conducting methodological research tasks and analyses of data user needs under the Generic Clearance of Survey Improvement Projects (OMB #3145-0174), as needed.

The primary effort will involve a follow-up to the SED Institution Exit Interview Review (as described above). The follow-up study will collect information about the ways in which SED institutions use (or might use) information from graduate student exit surveys, and the relative utility of selected exit survey items to different types of data users. The preliminary research plan includes conducting focus groups with graduate school deans, graduate department heads, other researchers/data users, and scholarship foundations. The focus groups will be followed by one-on-one interviews with individuals from the same four user groups to gather more in-depth information. The last component of the study will be a web survey that will be developed based on the results from the focus groups and interviews. It will be administered to the full population of graduate school deans (or other knowledgeable individuals) in the universe of SED institutions. The goal is to assess the potential value of expanding the SED to capture information on the graduate school experiences of doctorate recipients.

### ***B.5. Individuals Consulted***

The draft 2018 SED questionnaire was reviewed by the collaborating agencies in November 2016 (see Attachment 5 for the names of individuals consulted). See Attachment 2 for a list detailing the changes made to the 2018 SED questionnaire from the 2017 version and the rationale for those changes.



The RTI survey contractor staff who reviewed aspects of the survey design are listed in Attachment 5. In particular, the statistical experts from RTI associated with this survey are Ruth Heuer, Analysis and Reporting Task Leader (919-541-6457), and David Wilson, Statistical Task Leader (919-541-6990).

Additional individuals from both inside and outside NSF/NCSES who have been consulted on the statistical and methodological aspects of the design are also listed in Attachment 5. At NSF/NCSES, Lynn Milan, Project Officer for the SED (703-292-2275), and Samson Adeshiyan, NCSES Chief Statistician (703-292-7769), provide oversight for the survey.