

SECTION B: Collection of Information Employing Statistical Methods

Survey Data Collection Procedures Background

The SED questionnaire is distributed to new doctorate recipients by the Graduate Deans of the approximately 425 doctorate-granting institutions, and approximately 570 independent programs within those institutions, in the United States. The SED questionnaires (either paper or web) are filled out at the time the individuals complete all requirements for their doctoral degrees and are returned to NSF's survey contractor by the Graduate Dean. Because doctorates 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 around 92 percent.

When the completed paper survey questionnaires are received by the survey contractor, they are edited for completeness and consistency and then entered directly into the survey contractor's CADE program. Surveys received via the web survey mode do not need to be keyed and are edited mainly through a series of pre-programmed skip patterns and range checks. Errors which can clearly be remedied are corrected immediately; any questionnaire failing the edit for critical items will have a follow up MIL generated for the respondent. The MIL is sent either electronically or through regular mail and attempts to gather missing data on eight items (Bachelor's institution, year of Bachelor's, postgraduation location (state or country), birth date, citizenship status, race,/ethnicity, and gender), and for additional items (birth place, high school location, Master's year and Master's institution) if those items are also missing.

The survey contractor works with ICs to obtain contact information for students who have not submitted their SED questionnaires. An Address Roster is sent to ICs asking for the addresses of the nonresponders. Sometimes the IC's can provide other basic data items as well as the addresses. The survey contractor also utilizes Web-based locating sites to locate contacting information for nonresponders. A series of letters or emails is sent to any graduate who did not complete the survey through their graduate school, requesting their participation and containing a PIN/password for web access (see Attachment 9 for a sample letter).

Finally, any graduate who does not complete the SED through their graduate school and does not return a survey through the non-respondent mailing effort is given the opportunity to complete a slightly shortened version of the survey over the telephone. If, by survey close-out, an individual has not responded, public information from the commencement programs or other publicly accessible sources is used to construct a skeletal record on that individual. 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 earner. If a survey questionnaire is later received from a previous non-respondent, the skeletal record is replaced by the information provided by the respondent.

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. Because it is a census, no sampling is involved. 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 Department of Education. If so, the schools are asked to distribute survey questionnaires, or cooperate in the electronic distribution of the self-registration link, 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 research granting 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 added to the SED universe.

A high rate of response is essential for the SED to fulfill its role as a key part of the universe frame for longitudinal sample surveys, such as the Survey of Doctorate Recipients, and as the only reliable source of information on very small groups (racial/ethnic minorities, women, and persons with disabilities) in specialized fields of study at the Ph.D. 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 unacceptable. One reason many institutions participate in the survey is to receive complete information about all of their doctorate recipients in order to make comparisons with peer institutions. In addition, it is highly unlikely that the 573 graduate offices that voluntarily distribute the SED questionnaire could effectively carry out a sampling scheme such as handing out the questionnaire to every fifth doctoral candidate. This type of sampling would be even more difficult for the growing number of schools that use the web survey. In those cases, the school often refers the students to an online graduation checklist, where the SED is but one step in the graduation process. Finally, conducting the SED on a sample basis would produce poor estimates of small groups (in particular, 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 a much lower response rate because of difficulties in obtaining accurate addresses of doctorate recipients, particularly the foreign citizens who represent an ever growing proportion of the doctorates recipient universe each year. Such a technique would impose on the universities the additional burden of providing current addresses of new graduates, a somewhat ineffective process because experience with mailing surveys to new doctorates shows many addresses 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 annually grant almost 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. Matrix or item sampling – a

widely used technique in achievement testing – would not be feasible because the characteristic information is needed for each doctorate recipient for use in selecting the sample for the follow-up SDR. It would reduce the utility of the information to request, for example, sex, race, or field of degree information for some doctorate recipients and not for others. These characteristics are not evenly distributed across the doctorate population, and the extensive uses made of the data base rely on the completeness and accuracy of the information on doctorate recipients.

Therefore, sampling doctorates would decrease the utility of the data while increasing burden on the Graduate Schools which administer the survey and decrease the incentives for the institution to participate.

B.2. Survey Methodology

Because there is no sampling involved in the SED, there has traditionally been no weighting necessary. Basic information about non-responding individuals is obtained, where possible, from public records at their graduating institutions, graduation lists, etc. Both unit and item nonresponse are handled by including categories of “unknown” for all variables in tabulated results. The statistical and methodological experts associated with this survey are Dan Kasprzyk, Vice President, Center for Excellence in Survey Research at NORC (301-634-9396) and Mike Sinclair, Senior Fellow, Center for Excellence in Survey Research at NORC (301-634-9493), At NSF, Mark Fiegner, Project Officer for this survey (703-292-4622) and Stephen Cohen, NCSES Chief Statistician (703-292-7769), provide statistical oversight.

B.3. Methods to Maximize Response

The SED has enjoyed a high response rate during its existence, with an average of 92% completions over the past 30 years. It owes this high rate, in part, to the use of the data by the Graduate Deans, who go to extraordinary lengths to encourage participation on the part of their graduates. Each Graduate Dean receives a profile of their graduates, compared with other institutions in their Carnegie class, soon after the data are released each year. It is also due to extensive university outreach efforts on the part of the survey contractor, NORC at the University of Chicago, and National Science Foundation staff, and to the importance the universities themselves place on the data.

Throughout the data collection period, schools are constantly monitored for completion rates. Data on doctorates awarded on each commencement date are compared to data from the previous round in order to flag fluctuations in expected returns. Schools with late returns or reduced completion rates are individually contacted. Site visits, primarily to institutions with low response rates, by NSF staff and survey contractor staff are also critical to maintaining a high response rate to this survey. NORC’s electronic monitoring systems are particularly important to these efforts, as each institution’s graduation dates or SED submission dates can vary from monthly to annual.

In addition to the broad efforts to maintain high completion rates, targeted efforts to prompt for missing surveys and critical items are also key. The survey contractor works with ICs and also utilizes Web-based locating sites to contact students by mail and e-mail for missing surveys or items. 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. A series of letters is sent to any graduate who did not complete the survey through their graduate school, requesting their participation and containing PIN/passwords for web access plus paper questionnaires are sent to non-responding students. Additionally, a MIL is sent to any respondent who did not provide one of the critical items on the survey. Finally, any non-respondent who does not complete the SED through their graduate school and does not return a survey through the non-respondent mailing effort is given the opportunity to complete a slightly shortened version of the survey over the phone. Data received via the different modes are merged and checked to avoid duplicate requests going out to the various sources. 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 the SED data.

The response rates of institutions as well as the response rates to questionnaire items are evaluated annually. For example, the evaluation of the response rate for 2011 indicated that nearly half of the non-response was due to 20 institutions. Institutions with poor response rates were targeted for special letters or site visits by NSF or survey contractor staff and, to a large extent, these efforts have been successful in raising the response rates at institutions.

B.4. Testing of Procedures

The SED has undergone extensive review and testing of the questionnaire and the methods employed in conducting the survey in recent years. The changes made to the SED 2014 survey version are a result of many activities which have helped inform changes to instruments and procedures over time. The following major activities have been conducted since the previous OMB clearance submission (see Attachment 10.1 for a list of the methodological studies conducted over the past 15 years). The NSF project officer will be pleased to provide any of the documents referred to in this section or those referred to throughout the supporting statement.

Data Collection Related Tests

The accuracy of the data from the SED has been one of its strongest assets. An ongoing evaluation of the accuracy of coding, editing, and data entry processes is conducted. It consistently indicates that the error rate is very low (less than one percent). During data collection, the frequency distribution of variables is monitored on a continuous basis, so that emerging problems, such as high item non-response rates, can be identified early in the data collection phase and appropriate corrective measures implemented, if necessary. Additional quality control checks on the merger of paper and electronic questionnaires as well as the merger of missing information into the master data base are also ongoing. The survey questionnaires are constantly compared with the universities' graduation lists and commencement programs to make sure that only those persons with earned research doctorates are included.

In April 2012, as part of the larger efforts to expand web data collection, the SED survey contractor conducted semi-structured interviews with ICs from 16 institutions to better understand why some either did not respond to the initial outreach effort or are currently unwilling to transition to the Web-SED, as well as why other institutions were willing to transition. The 16 institutions were drawn from three groups: institutions among the 50 largest not currently using the Web-SED (we interviewed ICs from all but one such institution); institutions currently using the Web-SED; and smaller institutions that had not yet been approached with the full details of the Web-SED transition program. These interviews yielded valuable information for improving the current Web Outreach Strategy, summarized in the following recommendations which are currently being implemented:

- Expand and tailor initial contact strategy. An additional copy of all Web-SED transition materials should be sent to the IC as well as the relevant dean. Although the IC is often key in making the decision to transition to the web mode, some deans do not forward materials to the IC. The timing of this contact should be tailored to each school's graduation cycle.
- Differentiate the Web-SED from the PDF survey. The difference between the Web-SED and the PDF version of the survey should be emphasized so the benefits of the web mode are abundantly clear and attractive to the IC. Eventually, the PDF survey option should be phased out in order to streamline operations.
- Redesign outreach materials. The interviews showed that once ICs understand the benefits of the Web-SED they want to transition. Future versions of the outreach materials should focus on these benefits and explain the survey completion monitor process, clarify the level of effort required of the IC for the Web-SED, provide testimonials from institutions that have successfully transitioned, emphasize that the Web-SED is "green" and that the SED project is committed to reducing paper usage, highlight and explain Web-SED data security procedures, and provide a one-page handout with survey completion instructions for respondents.

Survey Quality Tests and Research

Several tasks were completed since the last OMB package, including several 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 round:

- Review of systems, programming, and quality control data preparation processes with a goal of earlier release of the data;
- 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 release of the data.

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;

- Review of the IPEDS database and the IRS form to determine if any institutions currently participating in the SED are offering eligible degrees that are not currently being included;
- Discussion of possible improvements in the coding and editing processes to ensure faster data entry resulting in more timely follow-up with non-respondents;
- Consultation with data processing managers on issues of paper and electronic data handling and mergers;
- In-depth analysis of confidentiality issues, particularly of data products that will be publicly available;
- Coordination of items common to the SDR and SESTAT instruments (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;
- Item analysis for floor and ceiling effects;
- Review of all respondent comments for concerns over confidentiality or item improvements;
- Review of “other, please specify” information in consideration of expanding or changing answer options;
- Coordination of items common to the SDR and SESTAT instruments, including the race/ethnicity and disability (i.e., “specific functional limitation”) items (see section A.4).

Finally, the following tasks were conducted during the last OMB clearance cycle, and will be conducted periodically in the future:

- Detailed review of emerging and declining fields of study and alignment with the CIP (Classification of Instructional Programs);
- Review of the non-PhD doctorate degrees included in the SED to confirm that they are research degrees and thus eligible for the survey;
- Literature reviews on targeted topics, such as disclosure avoidance and other confidentiality issues, as well as an initial review of the accreditation requirements for academic institutions.

Proposed Tests and Research

Over the course of the proposed OMB cycle (April 2013 – December 2015), NSF anticipates conducting several methodological research tasks and analyses of data user needs, some involving focus groups, cognitive interviews, and/or workshops. The tasks associated with these research studies and user analyses will be conducted under the Generic Clearance of Survey Improvement Projects package.

One research study will involve a review of the institutional criteria for participation in the SED. Historically, NSF and the other Federal Sponsors have set two criteria to establish whether institutions are eligible to participate in the SED: 1) the institution must offer a research doctoral degree, and 2) the institution should be accredited by one of the six regional accreditation organizations. However, the SED does include a few examples of non-accredited institutions (e.g., Rockefeller University) and NSF faces issues of determining whether certain newer,

accredited institutions (e.g., virtual institutions) meet the requirements for research doctoral programs. In response, NSF will conduct an analysis of the current eligibility criteria in order to develop clear, consistent eligibility criteria and/or eligibility review process. Implementing the new criteria/process in place will help the NSF avoid potential charges of implicit or explicit bias. The SED eligibility analysis will consider the impact of any changes on the fielding, implementation, and reporting of the SED and SDR, while addressing: 1) accreditation as an indicator of quality, 2) the rigor and comprehensiveness of the accreditation agencies' standards, and 3) reasons why some institutions decide not to seek accreditation. This will include reviewing accreditation agency websites, interviewing agency representatives and a small sample of accredited and non-accredited institutions offering research doctoral programs, and reviewing how other nations view and implement accreditation. The results will be documented in a report and a policy brief that summarizes the major findings and issues, and which could be widely disseminated. As an ancillary step, a panel of experts from different types of institutions will be convened for a one-day meeting in Washington DC. The charge to the panel will be to: 1) develop objective SED-eligibility criteria such as a definition for a research doctoral program and the need for accreditation; and 2) outline how to address issues of reporting, should the new criteria change the universe of the SED in significant ways. All findings will inform a decision by NSF regarding the institutional eligibility criteria and review process.

Another methodological research project planned for the next clearance cycle is related to improving the accuracy of debt reports. For reporting on education-related debt, the SED does not ask respondents to provide actual values but, rather, to select a set of dollar intervals into which their undergraduate and graduate education-related debts fall. As a check on the quality of these reported data, NSF will examine whether respondents can actually enumerate these values. Using decomposition, respondents will be asked about their debt for their undergraduate, then graduate, and finally total postsecondary education. This experimental examination will not only have consequences for the SED but also broader implications for collecting financial information in other surveys. The experiment will use two versions of the survey: (V1) will not change the current survey items, but will add a final item that ask respondents to provide an estimated total debt from the combined sources; (V2) will ask respondents to enumerate values for undergraduate, graduate, and total debt owed, but will not provide ranges. Both versions will include an item asking respondents to indicate their degree of confidence in their answers for the total value. The Web-SED versions will also collect client-side paradata on response changes, elapsed time, and response order. NSF will use the responses to the degree of confidence question to examine the effect the ranges have on providing numeric values, how decomposition influences actual value estimation and confidence, and how these changes affect respondent burden. This research could uncover new ways to ask respondents to report dollar values.

NSF will continue its investigation of alternative methods of disclosure avoidance in order to increase the utility and usability of data tables based on SED data. The current means of disclosure treatment in SED tabular data, cell suppression in tabulation (cs-tabulation), can have a severe impact on the reporting of doctorates awarded to underrepresented race/ethnicity groups and women, to the extent that information released on earned doctorates by fine field of study may be very limited. This is particularly undesirable for policy analysts working toward the goal of equal opportunity. One disclosure avoidance technique under consideration would replace

non-informative ‘D’ symbols with more informative, but still non-disclosive, model-based estimates for suppressed values.

NSF will continue the qualitative investigations of the communication of confidentiality assurances to institutions, data users, and survey respondents. This research will focus on how doctoral students, Graduate Deans, Institutional Contacts (IC), and Institutional Researchers understand the SED confidentiality assurances and guidelines, and how this understanding affects institutional and individual participation in the SED and the perceived utility of SED data. SED will conduct a series of focus groups and/or interviews with individuals from each of the previously mentioned stakeholder groups, and a sample of respondent comments will also be analyzed for confidentiality-related content. Findings from the focus groups and interviews will be supplemented with a textual analysis of the SED’s confidentiality-related material that is given to potential respondents and doctorate-granting institutions.

To ensure that these proposed methodological studies draw upon the collective experience and judgment of both NCSES and NCES and help align the two sets of surveys, where appropriate, NCSES commits to collaborating with NCES on the design of these studies and the evaluation of the results. For example, in the institutional eligibility study, the draft project report will be circulated to NCES for review and comment, and NCES staff will be invited to participate in the experts panel meeting held in Washington.

With respect to user needs analyses, NSF will conduct a study that will culminate in the design of data products specific to the needs of Graduate School Deans. The study will analyze the history of SED data requests made by deans, review the types of information posted on their graduate school websites, and conduct focus groups that bring Graduate Deans together to discuss their particular data needs. In another user needs analysis, NSF plans to research the utility and operational feasibility of providing SED data users with “early estimates” of selected SED data without compromising data quality. The results of the study will identify: 1) a preferred cut-off date for collecting the early-estimates data, 2) the trade-offs involved in not waiting to report the final data, 3) the best methods for addressing the situations involving missing data, and 4) scheduling implications.

Some revisions to the 2014 SED survey instrument are based on a NSF-sponsored review of the Web-SED conducted by Drs. Jolene Smyth and Kristin Olson from the University of Nebraska-Lincoln, which examined the SED web instrument in comparison to the SED hardcopy, SDR instrument and other relevant surveys. These analyses may continue during 2014 and 2015 rounds, and may include a methodology study to test the efficacy of the recommendations for Web-SED survey and paper questionnaire redesign. Interviews may be conducted with respondents to gauge their reaction to these changes, their reaction to the “Field of Study” lists on the paper survey versus the web survey, and other possible mode effects.

The draft SED 2014 questionnaire was first reviewed in December 2012, and the final questionnaire changes were reviewed and approved by the sponsors in January 2013. (See Attachment 5 for the list of persons who were consulted or who reviewed the questionnaire.) See Attachment 2 for a list detailing the changes made to the SED 2014 questionnaire from the 2011 version and the rationales for those changes.

B.5. Individuals Consulted

NORC at the University of Chicago is the organization contracted to collect and analyze the SED data for the 2014-2015 survey rounds. Staff from NORC who have consulted on the aspects of the design are listed in Attachment 5.

Additional individuals both inside and outside of NSF who have consulted on the statistical and methodological aspects of the design are also listed in Attachment 5.