

### ***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 550 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.

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 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. 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 experts associated with this survey are Colm O’Muirheartaigh, Vice President of Statistics and Methodology at NORC (312-759-4017) and Rachel Harter, Senior Statistician on the project at NORC (312-759-4058). At NSF, Mark Fiegenger, Project Officer for this survey (703-292-4622) and Stephen Cohen, NCSES Chief Statistician (703-292-7769), will 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, 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 is sent to ICs who can sometimes provide basic items 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 Missing Information Letter (MIL) is sent to any respondent who did not provide one of the “critical items” on the survey. This letter asks for the missing item and provides both a return envelope and a special e-mail address where the response can be sent. 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 2009 indicated that over half of the non-response was due to 29 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, and there has been extensive outreach to learn about the needs of SED data users. The changes made to the SED 2012 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.

## ***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;

- Extensive literature reviews on targeted topics, such as disclosure avoidance and other confidentiality issues, as well as the history and contemporary purpose of the Doctorate of Education (Ed.D).

### ***Research of SED Data Needs and Uses***

NCSES conducted a series of eight outreach meetings in fall of 2008 to learn about the specific data needs and uses of institutions, associations, and organizations that make extensive use of the SED's race/ethnicity and gender data. The meetings provided important input to NCSES's efforts to redesign the statistical tables that report this information. During this same period NCSES also conducted a web survey of a sample of SED data users – including the deans of all institutions participating in the SED – to gather similar information about the data needs and uses of different segments of the SED data user community.

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

Over the course of the proposed OMB cycle (April 2011 – April 2013), the SED anticipates conducting multiple methodological tasks involving focus groups and/or cognitive interviews with potential or already existing SED respondents. One set of interviews would involve the web survey. Dr. Don Dillman and colleagues have begun an expert review of the SED web survey and recently provided recommendations for changes to the design of individual questions. Dr. Dillman and colleagues will continue their expert review and may have further question changes as well as recommendations about the web survey methodology and its administration as a whole. 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, or other possible mode effects. The SED also anticipates conducting a web survey of all the ICs from participating institutions. This short survey would collect information on the specific practices involving the conduct of the survey at the institutions in the effort to identify new technologies, practices, or trends that impact the SED. The goal of this survey would be to identify areas where the SED can better support the ICs and adjust practices to meet the changing needs of the graduate schools. These methodological tasks will be conducted under the Generic Clearance of Survey Improvement Projects package.

Over the next two years the SED expects to conduct focus groups and workshops with representatives of different segments of the SED data user community. The purpose of the focus group meetings will be to raise emerging issues that may shape doctoral education in the future – issues that future SED data users will need to be informed about – and identify plausible metrics capable of tracking those issues. The data user workshops are intended to uncover the problems data users are having with the design of SED data tables or reports, and with the validity of particular SED data elements (with respect to how data users are using the data elements). The outputs of these focus group meetings and workshops will help inform NCSES decisions about new survey items that should be added to the SED, and the re-design of tables, reports, and existing survey items. These tasks associated with data user needs analyses will be conducted under the Generic Clearance of Survey Improvement Projects package.

The SED also anticipates conducting methodological studies that will not affect respondent or institutional burden. In 2011, NORC (at the request of NSF) will conduct research to test the

feasibility of using imputation to fill in missing critical items for non-respondents. This research will compare data tables using the available 2008 data and tables created using imputed data to look for possible impacts on data quality and published reports if imputation were introduced. At the request of several institutions, the SED will also explore the possibility of creating a transition page to come at the end of the SED web-survey that would allow respondents to link directly to their school's web-based exit survey.

The draft SED questionnaire was reviewed by Federal sponsors in November 2010, and the final questionnaire was reviewed and then approved by the sponsors in January 2011. (See Attachment 5 for the list of persons who were consulted or who reviewed the questionnaire.) See Attachment 2 for a list detailing changes made to the SED 2012 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 2012-2013 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.