

Attachment 6: Tests of Procedures Used (references in Section B.4)

6.1 Identifying the Intended Navigational Path of an
Establishment Survey

6.2 GSS Usability and Cognitive Interviews

6.3 Summary of GSS User Meetings

6.4 Pilot Studies

6.5 Process for Determining GSS-Relevant CIP Codes

Attachment 6.1

Identifying the Intended Navigational Path of an Establishment Survey

Cleo Redline¹

National Science Foundation
4201 Wilson Boulevard, Arlington, VA 22230, credline@nsf.gov

A desk stapler has sixteen parts, a household iron fifteen, the simple bathtub-shower combination twenty-three. You can't believe these simple objects have so many parts?

Donald A. Norman, "The Design of Everyday Things"

Introduction

We do not often see it described this way, but a self-administered survey instrument is really a physical object with many parts that need to work in unison for the express purpose of collecting information. These parts can be viewed along a continuum from the micro to the macro. At the micro level, there is color and brightness, shape and location of information. At a more macro level, there are the questions, instructions, definitions, and response categories. At an even more macro perspective, however, there is the entire mailing package—the outgoing envelope, cover letter, questionnaire, mailback envelope, possibly brochures or anything else that might be sent to respondents to aid and persuade them to answer and return the survey.

Making a survey instrument work, at both the micro and macro level, such that the parts are transparent to the respondents and the respondents can accurately and efficiently operate the survey instrument (defined as performing the tasks in the order respondents are instructed to perform them) is a monumental undertaking. This is especially true in the case of an establishment survey such as the Survey of Graduate Students and Postdoctorates in Science and Engineering (GSS) because of the large number of complicated parts that respondents are required to handle.

The objectives of this paper are: (1) to synthesize the literature that informed efforts to evaluate and improve the GSS, (2) to describe what were discovered to be the major parts of the GSS, (3) to show how the survey has been redesigned thus far and why, (4) and to extract the major principles that were used to redesign the survey.

Background

The literature is replete with design principles (e.g., Powell, 2002; Dillman, 2000; Couper, 1994; Norman, 1990; Wright and Barnard, 1975), but as Tourangeau (2000) points out the principles are vague enough that applying them is still as much an art as it is a science. Tourangeau also says that for this same reason it may be difficult to test them empirically or to determine the gains from questionnaires that embody them. On the other hand, we have made strides that make it worth at least attempting to synthesize what we have learned to date and worth attempting to extend it to establishment surveys. Take, for instance, the following principle:

1. Use the visual elements of brightness, color, shape, and location in a consistent manner to define the desired navigational path for respondents to follow when answering the questionnaire (Jenkins and Dillman, 1997).

It has since been shown that brightness, color, shape and location are visual elements of graphic language, and that graphic language is only one of the languages that comprises a questionnaire. Additional languages include the: (1) symbolic language, (2) numeric, and (3) verbal (Redline and Dillman, 2002). It is also now clear that 'in an effective manner,' not 'in a consistent manner' is the overarching goal really; consistency is only one means to this end. Finally, it has also become clear that it is necessary, but not sufficient, to think in terms of the questionnaire only. As Jenkins (1997) points out, improving the navigational qualities of a questionnaire requires paying attention to the entire mailing package. However, even that may not

¹ The views expressed in this paper are those of the author and do not necessarily reflect the views of the National Science Foundation. The author would like to thank her colleagues, Jeri Mulrow and Emilda Rivers, for their assistance and ingenuity throughout this project, and Scott Crawford (from Survey Sciences Group, LLC) for his advice and programming.

be sufficient, because as was discovered while redesigning the GSS, respondents may toggle between two modes: paper and Web. Thus, the following more encompassing and overarching revised principle emerges:

2. Use the *verbal, numeric, symbolic, and graphic languages* (i.e., visual elements of brightness, color, shape, and location) in a **consistent** and *effective* manner to define the intended navigational path for respondents to follow when answering the *survey (mailing package and/or Web questionnaire)*.

Identifying the Intended Navigational Path

A key phrase in the above design principle, a phrase that has remained stable from the first iteration of the principle to the second is “define the intended navigational path.” By definition, the navigational path is under the control of the interviewer in an interviewer-administered survey, whereas this path is under the control of the respondent in a self-administered survey. Four important tasks that respondents must handle in a self-administered survey that would be under the control of the interviewer in an interviewer-administered survey are: 1) starting correctly; (2) moving around; (3) reading and following instructions, definitions, or directions and; (4) answering in terms of the correct reporting unit. The addition of these tasks in a self-administered instrument translates into greater complexity for respondents, which in turn translates into greater respondent burden and potential for error. These tasks are not necessarily mutually exclusive, but are separated here for the sake of discussion.

Starting Correctly. The interaction that occurs within the first few seconds of a survey has long been thought critical to cooperation rates (Groves et al., 2005). According to sociolinguistics, one of the three key components of an interview is its opening interaction because this interaction sets the stage for the interview as a social encounter (Pan, 2003). The other two components are obtaining/giving information and maintaining the interaction. In cognitive interviews with three variants of a decennial census questionnaire (with approximately 20 respondents per variant), there was a large difference in how well respondents began the questionnaires. In one case, 4 percent of the respondents started by looking at the back page of the questionnaire, whereas 73 percent started by looking at the back page of another (Dillman et al., 1996). Respondents were asked in a debriefing at the end of the interview to recommend which questionnaire they suggested be used in Census 2000. One of the largest reasons they provided for their choice was the ease with which they were able to start the questionnaire. Forty-five percent favored the variant that allowed them to get started correctly compared to 22 percent for the one that did not (Redline, 1997). In the best-case scenario, a delay in starting correctly, however small or recoverable the delay may be, is a waste of respondents’ valuable time. In the worse case scenario, it will result in unit nonresponse. Numerous studies in Web surveys have demonstrated respondents’ inability to start correctly as a result of being provided ambiguous IDs or passwords, some of which have clearly been shown to lead to unit non-response (Fox et al, 2003; Couper et al, 2001).

Moving Around. Schwartz is generally credited with demonstrating that a survey adheres to the tenets of conversation (e.g., 1997), as put forth by Grice (1975). Because of its visual and physical nature, however, a self-administered survey is more like having that conversation while driving. Powell (2002) discusses many issues relevant to Website use: for example, site organization structures, navigational theory, navigational bars and the advantages and disadvantages of their possible placement, and the design of buttons. This information is discussed, however, in the context of Web sites, whose users have a different goal than respondents to a survey. Web site users tend to be searching for information, whereas with surveys, respondents are answering queries. As Fox et al (2003) have noted, there is less directly applicable information regarding the design of Web surveys. A finding for which evidence does seem to be accumulating, however, is that moving from screen to screen (or by extension, from page to page) leads to a disruption in visual continuity (Mockovich, 2005), which has been shown to lead to response errors in paper questionnaires (Redline et al., 2005; Featherston and Moy, 1990). Thus, moving from screen to screen or page to page (or even from the bottom of one column to the top of the next one) appears to be more susceptible to error than many would have us believe.

Reading and Following Instructions/Definitions/Directions. In a self-administered survey, respondents often need to read instructions, definitions, and directions critical to their understanding the questions and items, but they do not for what appear to be a number of, not necessarily mutually exclusive, reasons. One is because the information is often *located* physically out of sight (Redline et al., 2005; Tourangeau et al., 2003). Information may be far from where the respondent’s attention is presently. It may be in a separate booklet, for instance, or behind a link. Or, as has been demonstrated with skip instructions, it may be simply outside the respondent’s immediate attention (or foveal view in eye-movement terminology) (Redline and Lankford, 2001). Or, as has been shown with edit messages in Web surveys or branching instructions in paper questionnaires, it may be *located* after a disruption to visual continuity (Mockovich, 2005; Redline et al., 2005). A second possibility, however, is that the information is physically within sight, but is *located* among such a dense amount of information as to be rendered undetectable (Redline, 2003). A third possibility is that other aspects of the information’s graphic presentation (e.g., its color or size) give respondents the impression that the information is optional (Redline, 1997). A good example of this

comes from one of the Census 2000 test questionnaires in which the “Getting Started” information on the inside cover of the questionnaire was designed to take the place of the cover letter in the mailing package. This design failed, however, because the getting started information looked like unimportant background information. It was composed of blue lettering, while everything else was in black, and it was put outside the gold background area in what looked like unimportant floating white space.

Reasons endogenous to respondents, however, may compound the exogenous reasons provided above. That is, respondents may not perceive a need for the information (Redline et al, 2005; Frohlich, 1986). Or they may not have the cognitive ability to successfully process all of the information they are provided, or for various other reasons, they may lack the motivation to do so (Krosnick, 1991).

In terms of Principle 2 above, the numeric, symbolic and graphic languages need to unambiguously indicate what to read and the importance of reading it, and the verbal language needs to attract respondent’s attention, maintain their attention, and convey accurate meaning—not an easy undertaking. Furthermore, it is important that the methods used to identify respondent errors in self-administered surveys allow us to distinguish between respondents’ not reading information, and the reasons for their not reading it, as opposed to reading it and not comprehending it. We need to correct errors that arise as a result of both, but the solutions may differ.

Identifying the Reporting Unit. In a household or establishment survey, one of the first tasks respondents must undertake is to understand about whom or what they are being asked to report. Research has shown that answering in terms of the wrong reporting unit may be the result of respondents not starting the questionnaire correctly or not reading the necessary instructions. However, research has also shown that respondents may not read instructions and still report correctly, or they may read and understand the instructions, but not agree with them; or in the case of an establishment survey, it is conceivable that their records will not be in keeping with the request. Thus, it is clear that identifying and correcting the problems that arise here is not easy.

An example of not identifying the reporting unit correctly comes from cognitive research with the Public School 1991-92 Field Test Questionnaire for the Schools and Staffing Survey (SASS) (Jenkins et al., 1992). The cover page of this questionnaire contained a very important instruction in the paragraphs on the left-hand side of the page that told respondents to “Please complete this questionnaire with information about the SCHOOL named on the label.” Only about half of the 20 respondents ever read this instruction and looked at SCHOOL named on the label. Not reading the instruction mattered when respondents had alternative definitions of the school for which they could report, and got the wrong impression from the rest of the questionnaire regarding the unit for which they should report. The tendency was to mistakenly report for the entire school system (K-12) when they were really only supposed to report for the portion of the school over which they had direct charge (e.g., the high school).

The population count question (typically referred to as the pop count question) in the decennial census questionnaire has demonstrated similar tendencies. In cognitive interviews with the three variants of the decennial census questionnaire mentioned earlier, 15 percent of the respondents left the pop count question blank in the first variant, whereas only 7 and 6 percent left this question blank in the second and third variant, respectively. A large-scale mailout/mailback test of the three variants verified the fact that respondents tended to overlook the pop count question in the first variant of the questionnaire (Leslie, 1997). In hindsight, the pop count question appears like a mass of instructions in the left-hand corner relative to the name of Person 1 question, especially in the first variant of the questionnaire. These ‘instructions’ appear either unimportant or cognitively demanding in comparison to the much more salient and easy-to-answer Person 1 question.

The instructions in the pop count question are referred to as the residency rule instructions. Importantly, vignette research into the residency rule instructions has shown that respondents’ definitions of who should be reported as living or staying in their household is influenced by the residency rule instructions in some cases and not others. Gerber et al. (1996) refer to the cases in which the instructions have no effect as intuitive. For example, respondents do not need to be reminded to include permanent household members that are temporarily away. In contrast, the rules are necessary in counter intuitive situations to counter respondents’ erroneous preconceived definitions of who should be included and who should not. For example, according to census rules, someone’s mother who has been placed in a nursing home on a trial basis is supposed to be reported at the nursing home. A vignette regarding this situation showed the largest percentage gain (27 percent) in correct answers between the ‘with instruction’ and ‘without instruction’ conditions, demonstrating that the instruction helped.

Tourangeau et al. (2003) examined the complex everyday concepts of residence and disability and found similar results--that participants were better at classifying vignettes that closely matched a definition (*central* instances) than ones that only partly matched it (*peripheral* instances).

Methodology

This knowledge informed efforts to redesign the GSS. The GSS collects data from all U.S. institutions offering graduate programs in science, engineering, or selected health-related fields. Data published from the survey include graduate student enrollment by race, ethnicity, by gender, by type of financial support, by discipline, and by full-time or by part-time status. Data are also published on the number of postdoctorates and non-faculty researchers by gender, citizenship and by discipline. The GSS is a visually administered survey (with both a paper and Web component).

The redesign efforts consisted of conducting a round of cognitive/usability interviews with the original GSS, redesigning the survey, and conducting a second round of interviews with a redesigned portion of the survey. Twelve in-depth cognitive/usability interviews were conducted in the DC Metropolitan area and four in Florida in 2002 to learn more about how respondents navigated through and understood the original GSS. Interviews were conducted in the respondents' offices using the concurrent 'think aloud' interviewing method. Respondents were handed the paper mailing package and asked to report their thoughts aloud as they went about the process of actually handling the mailing package and responding to the survey questions. The mailing package provided instructions for accessing the Web survey. Respondents were encouraged to answer the survey by whatever mode they would normally answer it, and to retrieve the appropriate records to answer the survey. Being as interested in what respondents read as well as how well they understood what they read, respondents were asked to read aloud in addition to thinking aloud. When respondents fell silent or could not be understood, neutral, think-aloud probes were used as prescribed by Ericson and Simon (1980). Specific debriefing questions were held until the respondent had completed the think-aloud portion of the interview. In hindsight, the interviewing method was an attempt to use a method that would later be described by Hak et al. (2004) and discussed by Willis (2004, pg. 265). This method was designed to evaluate the cognitive aspects of answering the questions as well as the usability aspects of operating a visually administered questionnaire (either paper or Web). Interviews were both audio and video-recorded with the respondent's permission. The typical time for completion of the interview was two hours.

NSF contracted to have the survey redesigned in 2003-2004. Both the original and revised versions of the Web surveys were created in Cold Fusion (with possibly some Java script added) and were developed to run in Sequel Server. The paper version of the survey was created in PageMaker.

NSF contracted to have a second round of cognitive/usability interviews conducted in 2005 with a redesigned portion of the GSS. Eight schools were recruited from Chicago and Connecticut, respectively. However, due to cancellations, 5 interviews were actually conducted in Chicago and 7 interviews in Connecticut. The same interviewing method was used in the second round of interviews as had been used in the first, except this time around, respondents were sent an email an hour in advance of the interview, providing them with a link to the Web survey. As before, respondents were handed the paper mailing package at the beginning of the interview. Rather than asking respondents to complete the entire survey, they were asked to focus on a portion of the survey. And in addition to the cognitive/usability portion of the interview, a *reconciliation* was conducted at the end of the interview. That is, in advance of the interview, differences between the information provided in last year's GSS and information gathered from the institution's Website were discussed with respondents at end of the interview. The entire interview took about an hour. (Refer to Abt Associates, Inc., 2005, for a full reporting of this research.)

Results and Discussion

Identifying the Intended Navigational Path

One of the most important findings to come out of the first round of cognitive/usability interviews with the original survey was that the GSS is really composed of three separate parts or instruments. Part 1 is meant to elicit a list of departments in science, engineering, or selected health fields from survey coordinators, that is, to create the departmental frame. Part 2 is forwarded to a respondent to collect the data at the departmental level. Part 3 is a database management tool meant to assist survey coordinators monitor (either their own or the departments') data collection process. Furthermore, two of these parts were administered both by paper and by Web. Thus, it became evident that there were five parts or instruments, not just one. It also became evident that the overarching problem with the survey was that respondents were being asked to carry out many

tasks (refer to Table 1 for a listing of the tasks) and that the overall design of the survey did not highlight and reduce competition for respondent's attention to one manageable task at a time—in other words, it did not help respondents navigate through the information correctly, beginning with getting started correctly.

Another important finding to come out of the cognitive/usability interviews with the original survey was that although the majority of respondents submitted their survey by Web (for instance, about 75 percent of the respondents submitted their survey by Web at the time of the interviews), many of the respondents in the cognitive/usability interviews actually responded to the paper version of the survey, and simply transferred this information to the Web. In these cases, the true interface between the survey questions and that of the respondents was the paper questionnaire, with the Web survey serving as a data dissemination tool.

Signaling the Intended Navigational Path

Starting Correctly. In the original design of the paper questionnaire, Part 1 (listing science, engineering, and selected health-related departments) was downplayed and Part 2 (the questionnaire) was highlighted. Part 1 was printed portrait-style on standard white loose-leaf sheets of 8.5" x 11" paper stapled together in the upper left-hand corner. Its cover page was laden with instructions (full of text) that gave no visual clue that a task lay beneath (see Figure 1), that of listing the departmental frame (See Figure 2). Part 1 competed with the many other standard white mailing pieces that were included in the mailing package that were also text laden, similar to the way the residency rules in the pop count question of the decennial census questionnaire competed with the name question or the instruction on the cover of the Schools and Staffing Survey competed with the first question on that survey. As a result, a critical component of the GSS, the development of its frame, got less attention in the first round of interviews than was later assumed necessary to ensure correct reporting. In contrast, Part 2 was printed on colored paper stock in the form of a booklet and drew respondents' attention (see Figure 3). The Web version of the survey, on the other hand, was problematic for two different reasons: (1) it had an entirely different look and feel from that of the paper and (2) information on its getting started pages (accessing the Web survey and home page) competed for the respondents' attention. These problems will be described in greater detail next, and the solutions that were developed.

First, a new architecture with a new look and feel were conceptualized to better express or afford the parts in both the paper and Web and a prototype of Part 1 was developed. Visual design principles were used to make the two parts look as though they were supposed to function as a unit. Both were printed landscape in booklet formats with colored backgrounds, beige in the case of Part 1 (refer to Figure 4), and blue, in the case of Part 2 (refer to Figure 5), with the survey's heading and navigational bar expressed in white print against a dark blue background that spanned the top and right hand side of both booklets. In this conceptualization, the two parts were joined visually by their many similarities (that is, parallel construction), but were also distinguishable by the color of their background and by their salutations. The beige survey was addressed to the survey coordinator, the blue to the survey respondent. Since respondents tended to fill out the paper version of the survey and simply transfer this information to the Web in the first round of cognitive/usability interviews, the same look and feel was applied to both the paper and Web versions of the survey.

Costs played a significant role in determining which part received which color. The beige questionnaire required three colors to print: beige for the background, dark blue for the navigational bar, and black print for the text. Because three colors are more expensive to print and fewer coordinator questionnaires are printed than departmental questionnaires, the coordinator questionnaire was assigned the beige color. The blue questionnaire required only two colors to print: different tints of blue for the background and navigational bar, and black print for the text. Because two colors are less expensive to print, this color combination was assigned to the departmental questionnaire. It is interesting to note that because introducing color on the Web does not cost anything, originally a highly multi-colored design was proposed for the Web (see Figure 6). One of the goals of the redesign was to maintain consistency between the paper and Web versions of the survey, so that respondents could effectively navigate back and forth between the two modes. For this reason, and because the unrestrained use of color competes for respondents' attention, color choices were restrained in the redesigned version of the survey (refer to Figures 4, 5, 12, and 13 for examples).

Ensuring that the paper and Web surveys looked similar so that respondents would be able to navigate between them required a great deal of coordination and discussion between professionals with very different backgrounds—those who ordinarily worked on paper documents (editors) and who rarely, if ever, spoke to those who worked on the Web (programmers). These two spoke two very different languages, and rarely thought about how a change they made in one of the modes would affect

the other. It required a great deal of effort to ensure that differences between the two modes were introduced only after careful considerations on both sides.

Suggesting that Web surveys look similar to their paper counterparts is contrary to conventional advice now. The opposite advice, suggesting that Web surveys be different from their paper counterparts may have come about for two reasons: one is when paper questionnaires are badly designed, it is unreasonable to repeat what was a bad design to begin with. The second is the assumption that respondents answer by one mode or the other. If the GSS is any indication, respondents to establishment surveys may be answering the paper and submitting by Web. Thus, establishment surveys may need to be careful not to think of the two modes in isolation, but to ensure that they operate together.

To save space, only screens from the Web survey are shown and discussed in the remainder of this paper. It should be noted, however, that parallel pages from the paper survey tend to be identical, except for differences in functionality (differences due to buttons, links, etc.).

In the original Web survey, the welcome screen presented respondents with two links, one for the survey coordinators (school login screen) and another for the departmental respondents (department login screen). Refer to Figure 7. Respondents were often confused between these two links and chose the wrong one. The instruction preceding these links, "To login, please click School Login Screen or Department Login Screen" did not provide respondents with any additional information over what they knew already, which violated the conversational maxim of quantity (Grice, 1975). Furthermore, the login information, which is obviously the point of this screen, competed with the rest of the information on this page because it was placed in the same size and boldness as the rest of the information. Highlighted areas, such as the forget-your-password question beneath the login links, are likely to distract respondents' attention from the primary goal of logging in. Beyond the password section, the "Introduction" and "Why These Data Are So Important" sections are meant to persuade respondents' to complete the survey. These sections are either going to compete with respondents logging in or be totally ignored, neither of which serves a fruitful purpose at this point in time. This information should be processed, in other words presented, before reaching the login stage, perhaps in the cover letter.

Getting to the first question of the Web survey (the contact item) is twice as long in the original survey than the redesigned version. In the original, respondents must first respond to the login screen (Figure 8), a quick contact screen (Figure 9), and a main menu screen (Figure 10) before reaching the contact screen (Figure 11). On the main menu screen, they must select the number "1" from the middle of the page and from among many links and other pieces of information. Most of this information is premature at this point in time, violating the conversational maxim of relevance (Grice, 1975). The status legend, the upload and download your data links are not necessary until respondents get to the point where they are entering data, much later in the process. Furthermore, the navigational bars, the status legend, and other links are taking up valuable real estate: the top and left of the screen. Eye-movement research has demonstrated that all else being equal, the top and left of a page are more attended to than the bottom and right (Brandt, 1945). Thus, placing the navigational bar and other information at the top and left makes perfect sense if the purpose is to facilitate their use. However, since the primary purpose of a self-administered survey is to read and answer questions, it makes more sense to place the questions in the top and left-hand side of the page, and to relegate the navigational bar to a supporting role on the right.

In the redesigned survey, respondents are provided with a url, which automatically takes them to the correct login screen (see Figure 12). Furthermore, the screen they are taken to is highly simplified in comparison to the original. It highlights the task of logging in and rightfully reduces the original forget-your-password question to a supportive role (If you have forgotten your password...). It does this by boldly welcoming respondents and asking them to enter their User ID and password. The 'Welcome' salutation was purposely made big and bold and placed in the upper left-hand corner of the beige background, with nothing to its left or below it to distract respondents' attention from the primary task of logging in. The language here is consistent with the wording used in the cover letter, and with the sociolinguistic notion of this as an opening interaction. In other words, the computer jargon "to login" was purposely removed from the redesigned version and replaced with "to begin answering the online questionnaire."

The redesigned survey simplified and reduced getting to item 1 dramatically (refer to Figure 13). After logging in, the number 1 immediately presents itself in the upper left-hand corner of the page followed by the wording of the item. To further attract respondents' attention, the item wording was made bolder and larger than the rest of the information on the page, and clearly bolder and larger than it was in the original version of the survey.

The headings in the upper-left hand corner of the original survey were relocated because cognitive research has shown that respondents often read these headings in lieu of the items beneath them and draw the wrong interpretations as a result (Jenkins and Dillman, 1993). These findings are in direct opposition to conventional wisdom and research in the instructional text area, which has suggested that headings help readers understand text better (e.g., Hartley, 1981). However, one might reasonably conclude after reading this research more closely that *the information* conveyed by the heading is necessary, not necessarily the heading itself. Thus, an effort was made in the redesigned version to include all necessary information in the body of the item/questions. Besides helping readers understand the text better, headings serve another purpose, according to the instructional text research: they help readers to scan, select, or retrieve materials more easily. Hartley (1981) suggests that headings may be more usable when placed in the margin, although he also notes that this has not been subjected to research.

The redesigned survey makes use of these ideas. Rather than using the headings to supply pertinent information, they are used in the navigational bar in the right-hand margin to guide respondents through the survey. The navigational bar's design is predicated on the notion that items listed vertically signal greater independence than those listed horizontally (Jenkins and Dillman, 1997), along with the hypothesis that lists of action items are a commonly understood convention.

Finally, visual principles were used to indicate which item in the list is activated. This was accomplished by making the activated item look the same as the body of the screen (black text against a beige background). As a result, the selected item looks visually connected to the body of the screen, while the rest of the navigational bar is reduced to background information (white text on a dark blue background).

Moving Around. The cognitive interviews revealed that moving from screen to screen in the original survey was convoluted. In the original survey, respondents were presented with two buttons at the bottom of the contact information page: "Save and Refresh" and "Cancel," neither of which moved respondents to the next screen or question, as expected (see Figure 11). To move to the next screen they had to use the navigational bar to return to the main menu and choose where to go from there (which was not always evident), or choose where to go next from the navigational bar itself, or hit "Save and Refresh," followed by "Cancel," which then took them back to the main menu screen. Later screens presented even more complicated buttons from which to choose (e.g., as many as 8) in addition to the navigational bars at the top and bottom of the screens. This is an example of a hierarchical or 'tree structure' organization, with the home page serving as the 'root' of the 'tree structure' (Powell, 2002). In this organizational structure, backtracking is necessary to move forward. Furthermore, respondents have full control over the order in which they process information, which is not good if the point is to process information in a predetermined order. In this organizational structure, navigating through the survey is unnecessarily competing with actually answering the survey because it requires cognitive effort.

In the redesigned version of the survey, "Previous" and "Next" buttons were consistently placed at the bottom of the screen, not always in the same position on a screen, as is commonly recommended, but always the same distance beneath the last piece of information on the screen (see Figure 13). The point of placing them here is two-fold: (1) because respondents need to traverse this information to reach the buttons, thus they may be more likely to notice this information, or (2) if they do read the last piece of information, this is an efficient use of their eye-movements, and consequently, their time, for they simply need to continue down a little ways before reaching the 'next button.'

The 'Previous' and 'Next' buttons have been designed to serve as the primary means of navigating through the redesigned survey, with the navigational bar available in case respondents want to skip further ahead or back or exit the survey, and come back. This is an example of a linear-with-options organizational structure. This organization best mimics the paper survey in that it provides a structure for answering the survey in a pre-determined linear order, but it also allows respondents to freely move around, if they so choose. In this organizational structure, answering the survey is emphasized over navigating through the survey.

Reading Instructions/Definitions/Directions. The two modes of the original survey diverged greatly in the amount of instructions/definitions/directions provided to respondents in Item 2 "Maintain Your List of Departments." To accurately maintain their list of departments, survey coordinators need to know which departments should be included according to NSF's rules, and which should be excluded. As can be seen in Figure 14, however, the original Web version did not provide the survey coordinators with any guidance about which departments to include or exclude from the list, at least not without searching elsewhere for this information, and not one respondent in the first round of cognitive/usability interviews looked elsewhere for this information. The first instruction on the screen here "This screen allows you to maintain your list of departments" does not provide respondents with any additional information over what they were able to gather from the

heading already. Similarly, in the response matrix, there is a column labeled 'department name,' but no explanation as to what precisely that means.

Supplying an exhaustive list of rules, however, is not the answer either, especially if this list does not look as though it is part of the navigational path. The paper version of the survey (refer back to Figure 1) supplied the survey coordinators with so many instructions and in such a way (on white loose-leaf paper when the 'survey' itself was a colored booklet) that for all practical purposes it may have not been supplied. There were approximately 537 words on this page, and still not all the information that was needed is on this page. Some of it was buried elsewhere. Thus, in the redesigned version of the survey, an attempt was made to extract and reorganize the most critical of the eligibility rules and to place this information in the navigational path.

The redesigned version has approximately 303 words, a decrease from the original (see Figures 15 and 16 for the Web version). The first instruction on this screen is numbered '2,' and is placed in the same position on this page/screen as the number '1' had been in the preceding page/screen. The first instruction here attempts to convey to respondents that they are to answer three questions (A, B, and C) to update their list of departments in the table below, and questions that were only implied in the original were made explicit. In addition, columns that confused respondents in the cognitive/usability interviews, and for which no good purpose could be established, like the 'locked' column was removed. The locked column is in a spatially and temporally important position, the first column of the table, but it is not clear what its purpose is. And the second column was very confusing. No respondents in the cognitive/usability interviews referred to their departments by ID, but by their own names for their departments (e.g., Chemistry), which doesn't show up until the third column of the matrix.

In the redesigned version of the survey, department name was placed first, and what was an implied question in the original version of the survey was made explicit in the redesigned version: "In column A, what are the names of your school's science, engineering, and health-related graduate departments for 2004?" An attempt was made to link the question above the matrix to the correct column in the matrix by placing the prepositional phrase at the beginning of the question: "In column A". Furthermore, question A contains a very important piece of conceptual information that is not explicitly stated in the original version of the Web survey—the fact that this survey is supposed to collect information from health-related departments in addition to science and engineering.

Another piece of information that was missing from the original survey was the reference period. Schaeffer and Presser (2004) suggest that the reference period should be given at the beginning of a question (so that respondents do not construct their own before hearing the investigator's), and that respondents hear it before they formulate what they believe to be the intent of the question and that it be given in abbreviated form, and in a parallel location in subsequent questions. In this case that would have meant doubling up on prepositional phrases at the beginning of the sentence "In column A, in 2004," which seemed awkward. Furthermore, it would have been difficult placing the reference period in a parallel position in the rules that follow, so the reference period was repeatedly placed at the end of instructions instead. This is an example of the often-difficult choices survey designers must make. Besides creating a more complete, explicit question, the most relevant instructions pertaining to this question were divided by function (delete, remain or add) and listed in bulleted format under the question here.

A comparison of the table in the original version of the Web survey with the original paper version suggests that what is called "ID" in the Web version is referred to as "Department Code" in the paper version. This code is an effort on NSF's part to summarize departments into higher-order categories known as disciplines. The redesigned version did away with the code altogether, which has virtually no meaning for respondents. In the redesigned paper version of the survey, respondents are asked to refer to a list of disciplines, and in the Web, they are asked to select the discipline that best describes their department from a drop-down box.

One gets the impression from the heading in the original version of item 2 (maintain your list of departments) that this item is going to contain one task; however, closer inspection of the table reveals that respondents are actually being asked to conduct two: (1) to update their list of departments, and (2) to update the names of the departments' coordinators (refer to Figure 14). The redesigned version of the survey decomposed item 2 further. The first task (as shown in Figures 15 and 16) asks respondents to update their list of departments. The second task is not shown here, but is hinted at in the navigational bar. It invites departments to respond, and it is in this next task (or screen) that respondents are asked to update the names of the departments' coordinators.

Thus, an attempt was made to place a reasonable amount of information in the navigational path (e.g., making the listing form and eligibility rules look as though they are a part of the survey); reorganize the information in a way that is keeping with respondents' own schemas about the information (e.g., place department name first); connect physically disconnected but conceptually related information (e.g., use "in column A"); replace unfamiliar jargon with potentially more understandable terms (e.g., replace 'ID' or "Department Code" with "Discipline"); supply missing but important information (e.g., the reference period); and decompose the task into its constituent tasks (e.g., list departments then solicit information to aid in inviting departments to respond).

The second round of cognitive/usability interviews with the redesigned survey suggested that respondents did start more easily and correctly, navigate through the form with greater ease, and had a better understanding of what was being asked of them. Respondents deleted, added, and disentangled departments far more in the second round of interviews than in the first, suggesting that the changes made above were a step in the right direction. For example, one respondent noticed 5 dental departments that had been defunct for the past 5 years and deleted them during the course of the interview. (It is interesting to note that after participating in the cognitive/usability interview, she deleted 22 departments from the actual GSS.)

Defining the Reporting Unit. It was difficult to tell from the first round of interviews what were respondents' definitions for their reporting units. Instead of attending to Part 1 of the GSS in the first round of interviews, respondents had a tendency to accept whatever list they were provided with to start, and to focus on what they thought was the real purpose of the survey, the questions in Part 2. A few respondents, particularly those new to the survey, still had a tendency to view the pre-assigned list as correct in the second round of interviews; however, in general, a better understanding of respondents' definitions for their reporting unit was garnered from the second round of interviews for at least two reasons: (1) because an emphasis was placed on inducing respondents to read Part 1 in the redesigned version of the survey and (2) because the second round of interviews deliberately focused on this part of the survey.

Item 2 of Part 1 asks respondents to report a highly complex construct: *departments, programs or research centers with graduate students, postdoctorates, or non-faculty research staff in science, engineering, and selected health-related fields.* By most people's standards, the concept of postdoctorate alone is complicated, so the above construct is clearly highly complicated and multifaceted. As borne out by previous research regarding complex concepts (e.g., Gerber et al, 1996; Tourangeau et al., 2003), respondents' schemas were sometimes in keeping with NSF's definitions, and sometimes not. The second round of cognitive/usability interviews suggested that respondents had little difficulty reporting what Gerber et al. (1996) refers to as intuitive and Tourangeau et al. (2003) refer to as central instances. Examples of these in the GSS are physics and chemistry departments. However, respondents tended to have problems knowing whether to report what Gerber et al. (1996) refer to as ambiguous and Tourangeau et al. (2003) refer to as peripheral. A good example of this in the GSS is research centers. Finally, there were cases in the GSS that Gerber et al. (1996) refer to as counterintuitive, that is, respondents' definitions did not appear to agree with NSF's official definition of science, engineering, and health-related. An example of this was acupuncture.

Respondents not understanding NSF's technical definition for what to include and what to exclude, and not listing the institution's organizational elements correctly are likely to lead to *coverage error* at the second stage level. In general, respondents seemed to understand the redesigned version of the survey better in the second round of cognitive/usability interviews than in the first, although there were still problems. Respondents still came to the task with expectations (for example, believing that the pre-assigned list provided to them was accurate) and questions (wondering what the rules were about research centers and health fields) that need to be addressed in future renditions of this item.

Research shows that decomposing multifaceted or compound tasks into their constituent tasks and organizing information in ways that are in keeping with respondents' schemas or that cue their memories are helpful. Thus, we need to learn more about respondents' schemas in the future, and what cues will aid them access these schemas. It may prove beneficial to ask respondents to report their research centers first, then programs, and then departments (under the assumption that research centers are the least central to respondents' thinking about this issue, and therefore, the most likely to be forgotten or misunderstood). Also, the eligibility rules are now broken down by function (delete from the list, leave on the list, add to the list). Is this the best ordering of these functions? Might it be beneficial to separate these into separate questions, each referring to a separate response table? However, what is best if respondents are filling out the paper questionnaire, and transferring this information to the Web? Further research is needed to help us answer these difficult questions.

It should be noted that a pattern appears to be emerging from household and establishment surveys (Jenkins et al, 1992; Gerber et al., 1996), which when put together with the research of Tourangeau et al. (2003) suggests that respondents can

have multiple definitions for their reporting units and that they need to clearly understand what the survey is requesting (that is the survey's eligibility rules) in order to report correctly. The findings of the research seem to point in the same direction: household and establishment surveys, which have a global instruction buried at the beginning of the survey aimed at respondents' understanding the unit for which they are to report, may not be elicit accurate reports. These surveys may benefit from following the GSS' model of listing out their organizations' elements (e.g., members of a household; subsidiaries, affiliates, or branches of a business; or departments of an institution) first, followed by requesting data from/about those elements.

Redesign Principles. Many changes were made to the GSS to induce respondents to start it correctly and to better understand the tasks they were being asked to perform in the order they were being asked to perform them, with a focus on getting respondents to properly attend to the critical task of defining their reporting unit. The major principle used to guide these changes was provided at the beginning of this paper: use the verbal, numeric, symbolic, and graphic languages in an effective manner to define the intended navigational path for respondents to follow when answering a self-administered survey (Jenkins and Dillman, 1997). Table 2 attempts to delineate this principle into the sub-principles that were used to redesign the GSS. These sub-principles may provide even more concrete guidance for improving a survey, or they may serve as input for hypothesis testing under controlled, scientific conditions. Harking back to the major principle, these sub-principles attempt to define how to signal the navigational path more 'effectively.'

Conclusions and Future Research

Starting with the end in mind does not mean starting at the end. In the case of the GSS, the end is to elicit accurate counts of graduate students, postdoctorates, and non-faculty research staff in science, engineering, and health-related departments, programs, and research centers. The results of previous survey research suggest that to reach this goal, we need to start much further back. We need to help respondents: start the GSS correctly; move through the survey in a systematic order; read the questions, instructions, definitions, and directions in non-critical items so that by the time they reach a critical item, they are primed to read information which will aid them to accurately answer it with as little effort as possible. In other words, we need to identify and define the navigational path for respondents.

The first round of cognitive/usability interviews with the original survey confirmed the fact that the navigational path was not clear, and that respondents had a tendency to overlook Part 1 of the survey (and more specifically, Item 2 in Part 1). This item elicits the survey's departmental frame—or more specifically, a list of science, engineering, and health-related departments, programs, or research centers that have graduate students, postdoctorates, or non-faculty research staff. A second round of cognitive/usability interviews with the redesigned version of the survey suggested that respondents did start the redesigned survey more easily and correctly, navigate through it with greater ease, and had a better understanding of what was being asked of them in this item. Respondents deleted, added, and disentangled departments more in the second round of interviews than the first.

An important finding to come out of this research was that respondents often answered the paper version of the GSS and then simply transferred this information to the Web. Thus, the navigational path of the redesigned survey was in effect doubled—respondents not only needed to navigate *within* a mode, but *between* two modes. Contrary to conventional advice, it became imperative to maintain consistency between the two modes so that respondents could navigate back and forth between the two modes without getting lost.

Another important finding to come out of the redesign efforts was that in order to answer the GSS correctly, respondents must understand a very complex and multifaceted technical construct, that of *departments, programs or research centers with graduate students, postdoctorates, or non-faculty research staff in science, engineering, and selected health-related fields*. It is clear that more research needs to be conducted before this construct will be successfully conveyed to respondents.

Finally, one of the most important findings to come out of this research is that instructions provided at the beginning of household and establishment surveys, aimed at helping respondents' understand the unit for which they are to report, are often inadequate. In all likelihood, these surveys would benefit from creating an additional step (or part) at the beginning of the survey, in which respondents are asked to enumerate the organization's elements according to some rules (e.g., members of a household; subsidiaries, affiliates, or branches of a business). Thus, we end where we began: a self-administered survey instrument is really a physical object with many *parts* that need to work in concert for the express purpose of collecting information. It is clear that we need to continue to understand these *parts* ourselves, and to help respondents navigate through them and understand them as well.

References

- Abt Associates, Inc. 2005. NSF-GSS Final Report: Cognitive/Usability Testing of Module One in the Spring of 2005. Report Prepared for National Science Foundation, Arlington, VA 22230.
- Brandt, H.F. 1945. *The Psychology of Seeing*. New York: The Philosophical Library.
- Couper, M. 1994. What can CAI learn from HCI? Paper presented at the COPAFS Seminar on New Directions in Statistical Methodology. June 1994, Bethesda, MD.
- Couper, M., Traugott, M.W, and Lamias, M.J. 2001. Web Survey Design. *Public Opinion Quarterly*. 65:2.
- Dillman, D. 2000. *Mail and Internet Surveys: The Tailored Design Method*. John Wiley & Sons: New York.
- Dillman, D., Redline, C., Martin, B., and DeMaio, T. 1996. Cognitive and Motivational Properties of Three Proposed Decennial Census Forms. Report Prepared for the U.S. Bureau of the Census. Washington, D.C.
- Ericson, K.A., and Simon, H.A. 1980. Verbal Reports as Data. *Psychological Review*. 87: 215-251.
- Featherston, F. and Moy, L. 1990. Item Nonresponse in Mail Surveys. Paper Presented at the International Conference on Measurement Errors in Surveys, Tucson, Arizona.
- Fox, J.E., Mockovak, W., Fisher, S.K, and Rho, C. 2003. Usability Issues Associated with Converting Establishment Surveys to Web-Based Data Collection. Proceedings of the Federal Committee of Statistical Methodology. Washington, D.C.
- Frohlick, D. 1986. On the Organizations of Form-filling Behavior. *Information Design Journal*. 5:43-59.
- Gerber, E., Wellens, T.R., and Keeley, C. 1996. Who Lives Here? The Use of Vignettes in Household Roster Research. *Proceedings of the Section on Survey Research Methods*, (pp 962-967) Alexandria, VA: American Statistical Association
- Grice, H.P. 1975. Logic and Conversation. In P. Cole (ed.) *Syntax and Semantics*. Vol 3. (pp. 41-58) New York: Academic Press.
- Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., and Tourangeau, R., 2004. *Survey Methodology*. John Wiley & Sons: Hoboken, New Jersey.
- Hak, T., van der Veer, K. and Jansen, H. 2004. The Three-Step Test-Interveiw (TSTI): An observational instrument for pre-testing self-completion questionnaires. (ERIM Report ERS-2004-029-ORG). Rotterdam: Erasmus Research Institute of Management. Retrieved January 05, 2006 from hdl.handle.net/1765/1265.
- Hartley, J. 1981. Eighty ways of Improving Instructional Text. *IEEE Transactions on Professional Communication*, Vol. PC-24, No. 1, March 1981.
- Hunter, J., Redline, C., Morrison, R., Willimack, D., Sangster, R. 2005. Broadening the Horizons of Q-bank: Expanding the Framework to Encompass Self-Administered and Establishment Surveys. Paper prepared for presentation at the American Statistical Association, Minneapolis.
- Jenkins, C. and Dillman, D. 1997. Towards a Theory of Self-Administered Questionnaire Design. In L. Lyberg, P. Biemer, M. Collins, E. DeLeeuw, C. Dippo, N. Schwarz, & D. Trewin. (Eds). *Survey Measurement and Process Quality* (pp. 165-196). New York: Wiley.

- Jenkins, C. 1997. Improving the Navigational Qualities of the Decennial Census Short Form Requires Paying Attention to the Entire Mailing Package. *Proceedings of the Section on Survey Research Methods*, Alexandria, VA: American Statistical Association.
- Jenkins, C., Ciochetto, S. and Davis, W. 1992. Results of Cognitive Research on the Public School 1991-92 Field Test Questionnaire for the Schools and Staffing Survey (SASS-3A). Report Prepared for the National Center for Education Statistics. Washington, D.C.
- Krosnick, J.A. 1991. Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5:213-236.
- Leslie, T.F. 1997. Comparing Two Approaches to Questionnaire Design: Official Government Versus Public Information Design. *Proceedings of the Section on Survey Research Methods*, American Association of Political Opinion Research. (pp. 336-341) Alexandria, VA: American Statistical Association.
- Mockovich, B. 2005. An Evaluation of Different Design Options for Presenting Edit Messages in Web Forms. Paper Presented at FedCASIC, Bureau of Labor Statistics: Washington, DC.
- Norman, D. 1990. *The Design of Everyday Things*. Doubleday: New York.
- Pan, Y. 2003. The Role of Sociolinguistics in the Development and Conduct of Federal Surveys. *Proceedings of the Federal Committee of Statistical Methodology*. Washington, D.C.
- Powell, T.A. 2002. *Web Design: The Complete Reference*. Second Edition. McGraw Hill/Osborne: New York.
- Redline, C., Dillman, D., Carley-Baxter, C., Creecy, R. 2005. Factors that Influence the Reading and Comprehension of Branching Instructions in Self-Administered Questionnaires, *Allgemeines Statistisches Archiv*. 89, 21-38.
- Redline, C. 2003. Discussion of Tourangeau, Couper, and Conrad's paper, The Impact of the Visible. The Second Seminar on the Funding Opportunity in Survey Methodology. Research Subcommittee of the Federal Committee on Statistical Methodology. Washington, DC.
- Redline, C. and Dillman, D. 2002. The Influence of Alternative Visual Designs on Respondents' Performance with Branching Instructions in Self-Administered Questionnaires, in Groves, R., Dillman, D., Eltinge, E., and Little, R. (eds.) *Survey Nonresponse*. New York: John Wiley and Sons, Inc.
- Redline, C. and Lankford, C. 2001. Eye-movement Analysis: a New Tool for Evaluating the Design of Visually Administered Instruments (paper and Web), *Proceedings of the Section on Survey Research Methods*, Alexandria: VA: American Statistical Association.
- Schaeffer, N. C. and Presser, S. 2003. The Science of Asking Questions. *Annual Review of Sociology*, 29: 65-88.
- Schuman, H. and Presser, S. 1981. Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. San Diego: Academic Press.
- Schwartz, N. 1997. Questionnaire Design: The Rocky Road from Concepts to Answers, in Lyberg, L., Biemer, P., Collins, M. de Leeuw, E., Dippo, C., Schwarz, N., Trewin, D. (eds.) *Survey Measurement and Process Quality* New York: John Wiley & Sons, Inc.
- Tourangeau, R., Conrad, F., Arens, Z., Fricker, S., Sunghee, L., and Smith, E. 2003. Everyday Concepts and Classification Errors: Judgments of Disability and Residence. Paper presented at the American Association of Public Opinion Research, Phoenix, Arizona.

Tourangeau, R., Couper, M., and Conrad, F. 2003. The Impact of the Visible. Paper presented at the Second Seminar on the Funding Opportunity in Survey Methodology. Research Subcommittee of the Federal Committee on Statistical Methodology. Washington, DC. June, 2003.

Tourangeau, R., Rips, L., and Rasinski, K. 2000. *The Psychology of Survey Response*. Cambridge University Press: Cambridge.

Willis, G. 2005. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, California: Sage Publications.

Wright, P., and Barnard, P. 1975. Just Fill in This Form—A Review for Designers. *Applied Ergonomics*, 6:213-220.

Table 1. A Listing of the Self-Administered Tasks in the GSS by Type and Number of Respondents and Mode of Administration

Self-Administered Tasks	Respondent Type	Number of Respondents	Mode of Administration
1. Email cover letter Read email Link to Web survey, or go to mailing package, or both	Survey Coordinator	1	Web
2. Mailing package envelope Read envelope Open envelope Take out contents	Survey Coordinator	1	Paper
3. Questionnaire cover letter Read letter Link to Web survey or Go to Module 1 or both	Survey Coordinator	1	Paper
4. Part 1:	Survey Coordinator	1	Paper or Web
4.1. Review the contact information Read information Comprehend Retrieve information Act: respond			
4.2. Enumerate the sampling frame (science, engineering and selected health- related departments) Read information Comprehend Retrieve information Respond			
Distribute the questionnaire(s)/ invite departments to respond			
5. Part 2:	Survey Coordinator or Department Respondent	1 up to as many respondents as departments	Paper or Web
Read cover letter			
Read instructions and definitions			
Respond to department questionnaire Read items Comprehend Retrieve information Judgment Respond Return or submit dept. questionnaire			
6. Part 3:	Survey Coordinator	1	
View department progress			Web
Contact Departments about Corrections			Paper or Web
7. Return or Submit Survey			Paper or Web

Table 2. Design principles, divided by language type.

Graphic

- Location
 - Up and left of page/screen indicates the beginning, thus place all information in the order it will be used temporally, beginning with upper left
 - Place numbers to the left of text and text to the left of action items
 - Place response categories to the left of check boxes
 - Place check boxes to the left of skip instructions
 - Place navigational bar to far right of screen (because in essence it is a skip instruction)
 - Place definitions/instructions/error messages within close proximity of where they are to be acted upon
 - Bottom and right of page/screen indicates end, thus place next button last on screen just beneath the last piece of information to be read in the lower right side
 - Place conceptually related information horizontally (rows)
 - Place conceptually **unrelated** information vertically (columns)
- Brightness and color
 - Use high contrast and/or color coding
 - Use light background
 - Light beige background (20 percent saturation) for coordinators, light blue background (20 percent saturation) for departments
 - Use dark text, preferably black
 - Black text (100 percent saturation)
 - Use red error messages to distinguish them as different from other text
 - Use blue Web links because this is convention
 - Use size and boldness to convey hierarchy (i.e., establish order of importance).
 - Ensure that size and boldness do not conflict with location, but work in concert.
 - Use white for action items (radio buttons, text boxes, drop down boxes, navigational links, etc.)
 - Do not rely on color only to convey meaning
 - Minimize changes in color: less is more
- Shape
 - Use san-serif font
 - Minimize changes in font under the auspices that less is more

Symbolic

- Use symbols in addition to verbal messages
- Use familiar symbols
- Beware of ambiguity

Numeric

- Use numbers to logically order questions for respondents
- Beware of ambiguity (for example, using '1' more than once)

Verbal

- Use a conversational (social interaction) style, as opposed to computer jargon
- Present only necessary and relevant information (Maxims of Quantity and Relevance (Grice, 1975))
- Use short, active, affirmative sentences/questions (Maxim of Manner (Grice, 1975))
 - One clause
 - Active rather than passive
 - Affirmative rather than negative
 - Decompose questions: ask questions about one thing at a time
- Use familiar words (rather than unfamiliar words, numbers, or symbols) (Maxim of Manner (Grice, 1975))
- Beware of ambiguities (Maxim of Manner (Grice, 1975))

NATIONAL SCIENCE FOUNDATION AND NATIONAL INSTITUTES OF HEALTH
Survey of Graduate Students and Postdoctorates in Science and Engineering: Fall 2001

INSTITUTION: 999999-8 Test Institution
 Test School

INSTRUCTIONS FOR SURVEY COORDINATOR

1. Please review the list of departments/programs on the following pages. Add any newly formed graduate science or engineering (S&E) departments or programs in the spaces provided at the end of the list; note any name changes or reorganizations where they occur. Because of special interest at NSF and NIH, please separate the following: computer science from mathematics; all engineering fields by discipline (electrical, mechanical, etc.); and clinical psychology from all other psychology programs. Please submit separate survey Forms 812 for each such program.
2. Please cross out discontinued or non-S&E programs on the list, and indicate the last year that each discontinued program had students. However, programs that do not have graduate students or postdoctorates in Fall 2001, but may have them in the future, should NOT be crossed out; instead place a checkmark in column (6). Do NOT submit a Form 812 for such departments/programs.
3. The enclosed Crosswalk shows the 3-digit NSF code and comparable NCES discipline codes for each S&E field. Please review the first 3 digits of the codes in column (2) of your list to verify that each department/program has been assigned to the correct discipline. If you do not agree with the code in column (2), please indicate the Crosswalk code that most closely corresponds to your department/program.
4. (a) In MEDICAL SCHOOLS, departments/programs include the indented sections/specialties listed beneath them. These sections/specialties may be without organizational autonomy; however, because of special interest at NIH, please submit separate data for each section/specialty. In addition, please make sure that no duplicate data are reported for sections/specialties and their parent programs. (b) In GRADUATE SCHOOLS, departments include the indented specific degree-programs listed beneath them. Please submit separate Forms 812 for each specific degree-program.

5. Please distribute the enclosed Forms 812 (survey questionnaires) as soon as possible to all listed and newly formed departments/programs if data cannot be obtained from your centralized files. The Forms 812 provide all needed instructions and definitions. Please follow these guidelines to complete the Forms 812.
6. Please review the Forms 812 for completeness and consistency, and verify or supply the highest degree offered by each department/program. Please check this year's data against the previous year's data shown on the list, and indicate the validity of any large year-to-year changes in the "Comments" section below or on the individual Forms 812.
7. Please do NOT submit responses for any graduate departments in education, law, humanities, music, the arts, physical education, library sciences, and all other non-S&E fields.
8. Please return one copy of the Form 811 (this form) along with all completed Forms 812 by January 31, 2002, to:

QRC Division of Macro International Inc.
7315 Wisconsin Avenue, Suite 400W
Bethesda, MD 20814-3202

Please keep copies of all forms for your records.

9. In order to minimize followup procedures and make more timely statistics available, please submit as complete and as accurate a response as possible. If you have any questions about your response, please contact Jason Bauer of QRC at (301) 657-3077, ext. 322, or by email at jbauer@qrc.com. You may also contact Dr. Joan Burrelli at NSF's Division of Science Resources Statistics at (703) 292-7793, or by email at jburrell@nsf.gov.

SURVEY COORDINATOR CONTACT INFO:

David Greene
TELEPHONE: 301-657-3077
EMAIL: dgreene@qrc.com

SIGNATURE OF SURVEY COORDINATOR

COMMENTS:

DATE COMPLETED

Figure 1. Part 1's original cover page.

NATIONAL SCIENCE FOUNDATION AND NATIONAL INSTITUTES OF HEALTH
Survey of Graduate Students and Postdoctorates in Science and Engineering: Fall 2001

INSTITUTION: 999999-8

Test institution
Test School

LIST OF DEPARTMENTS AND PROGRAMS

2001 RESPONSE HISTORY (1)	DEPARTMENT CODE * (2)	S&E DEPARTMENTS OR PROGRAMS (3)	HIGHEST DEGREE GRANTED (4)	FALL 2000 FULL TIME^^ (5)	FALL 2000 PART TIME^^ (5)	FALL 2000 POST- DOCS^^ (5)	CHECK HERE IF NO STUDENTS AND NO POSTDOCTORATES (6)
Data	20201	Chemistry	Ph.D. or equivalent	90	4	20	
Data	40202	Mathematics	Ph.D. or equivalent	121	120	38	
Data	60301	Biology	Ph.D. or equivalent	110	76	1	
TOTAL, ALL DEPARTMENTS OR PROGRAMS:				321	200	59	

* The first three digits of the department code are a field of science and engineering code; a list of these codes is enclosed.
^^I = data were computer estimated; E = data were estimated

Note: This information is solicited under the authority of the National Science Foundation Act of 1950 as amended. All information will be used for statistical purposes only. Your response is entirely voluntary and your failure to provide some or all of the information will in no way adversely affect your institution.

LIST OF DEPARTMENTS AND PROGRAMS

ADDITIONAL DEPARTMENTS OR PROGRAMS FOR FALL 2001:

Please fill in columns (3)—S&E Departments or Programs, (4)—Departmental Degree Level, and if applicable, (6)—No Students and No Postdoctorates. Using the NSF Crosswalk provided, please indicate in column (2)—Department Code the NSF discipline code that you think is most appropriate, or provide a brief description of the department/program's major area of concentration.

DEPARTMENT CODE (2)	S&E DEPARTMENTS OR PROGRAMS (3)	DEPARTMENTAL DEGREE LEVEL (4)	CHECK HERE IF NO STUDENTS AND NO POSTDOCTORATES (6)
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____

DATA SOURCE

The data on the survey forms are derived from which of the following sources?

- a computerized central records system
- automated systems relying on departmental input
- hand-tabulated at the institutional level
- hand-tabulated at the departmental level
- a combination of the above sources

Figure 2. Part 1's original departmental listing page.

NATIONAL SCIENCE FOUNDATION
ARLINGTON, VA 22230

Dear Departmental Respondent:

The National Science Foundation (NSF) and the National Institutes of Health (NIH) request your participation in their annual Survey of Graduate Students and Postdoctorates in Science and Engineering. Your timely response will ensure that your department is accurately represented in data used by Federal and State decision-makers. In addition, we hope you will find the data useful for comparisons with other institutions in your State or region. Also, since industrial firms use these data in preparing for employment recruiting visits, your participation in the survey provides valuable benefits to your graduates.

Your survey data should be returned to NSF on or before January 31, 2002. This year the survey will continue to be offered online. The Web-based data collection system (<http://www.qrc.com/gssweb/gss.htm>) offers many advantages over the paper survey form. We encourage you to take advantage of the Web system. Departmental IDs and passwords to access the system are available from your school coordinator or from Mr. Jason Bauer of QRC Division of Macro International Inc. If you do not use the Web system, please return your completed data sheet to your survey coordinator before this date so that it can be transmitted to NSF in time to meet the survey deadline.

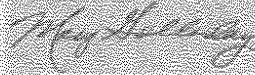
The data requested for the fall 2001 survey are exactly the same as those for the fall 2000 survey. When you return the survey Form 812 to your institution's survey coordinator, please keep a copy for your files. The instructions on the survey form and "How to Avoid Common Survey Errors" on page 7 should assist you in completing the survey. Because survey data with arithmetic errors or large trend inconsistencies will be returned for correction, verification, or explanation, please adhere to the guidelines. If you have any questions about the survey forms, please contact Mr. Bauer by telephone at (301) 657-3077, ext. 322, or by email at gss@qrc.com.

Please provide separate data sheets for (1) clinical psychology and (2) all other psychology departments or degree-programs, due to special interest in the clinical area. Similarly, please submit separate data for degree-programs in engineering, such as electrical engineering or mechanical engineering, and likewise, separate data for degree-programs in computer science and in mathematics.

Survey results are made available in a variety of modes. A two-page *Data Brief* and selected summary tables are published; all tables and reports are made available in various electronic formats on the Internet (<http://www.nsf.gov/sbe/srs/stats.htm>). These results are also entered in the Web-Based Computer-Aided Science Policy Analysis and Research (WebCASPAR) database system. The URL for WebCASPAR is <http://caspar.nsf.gov/webcaspar>.

Thank you for your efforts in providing timely and meaningful statistics to NSF and NIH.

Sincerely,



Mary J. Golladay, Ph.D.
Program Director
Human Resources Statistics Program
Division of Science Resources Statistics
National Science Foundation

Sincerely,



Walter T. Schaffer, Ph.D.
NIH Research Training Officer
Research Training Office
Office of the Director
National Institutes of Health

Enclosures: D

Figure 3. Part 2's original cover page.

NSF-NIH 2004 Survey of Graduate Students and Postdoctorates in Science and Engineering

Northwestern University
Medical School
FICE Code: 001739-5

For Survey Coordinator

Use this questionnaire as a guide while answering the online questionnaire:

1. Go to <http://gssredesign.us>.
2. Enter your user ID: **001739-5**.
3. Enter your password: **password**.

Alternatively, answer by paper only:

1. Complete and return the postcard inside this questionnaire.
2. Update the list of departments/programs.
3. Complete or distribute the enclosed blue questionnaires.
4. Return completed questionnaires in the enclosed envelope by January 31, 2005:

NSF-NIH Graduate Student Survey
ORC Macro
7315 Wisconsin Avenue, Suite 400W
Bethesda, MD 20814

OMB No. 3145-0174
Appr. Exp. 12/2006

Figure 4. Part 1's redesigned cover page.

2004 Survey of Graduate Students and Postdoctorates in Science and Engineering

Institution Name
School Name
FICE Code: xxxxxx-x

For Survey Respondent

1. Complete the Department or Program Data Sheet
2. Complete items 5-8 in the questionnaire
3. Return completed questionnaires in the enclosed envelope by January 31, 2005, to:

NSF-NIH Graduate Student Survey
ORC Macro
7315 Wisconsin Avenue, Suite 400W
Bethesda, MD 20814

Figure 5. Part 2's redesigned cover page.

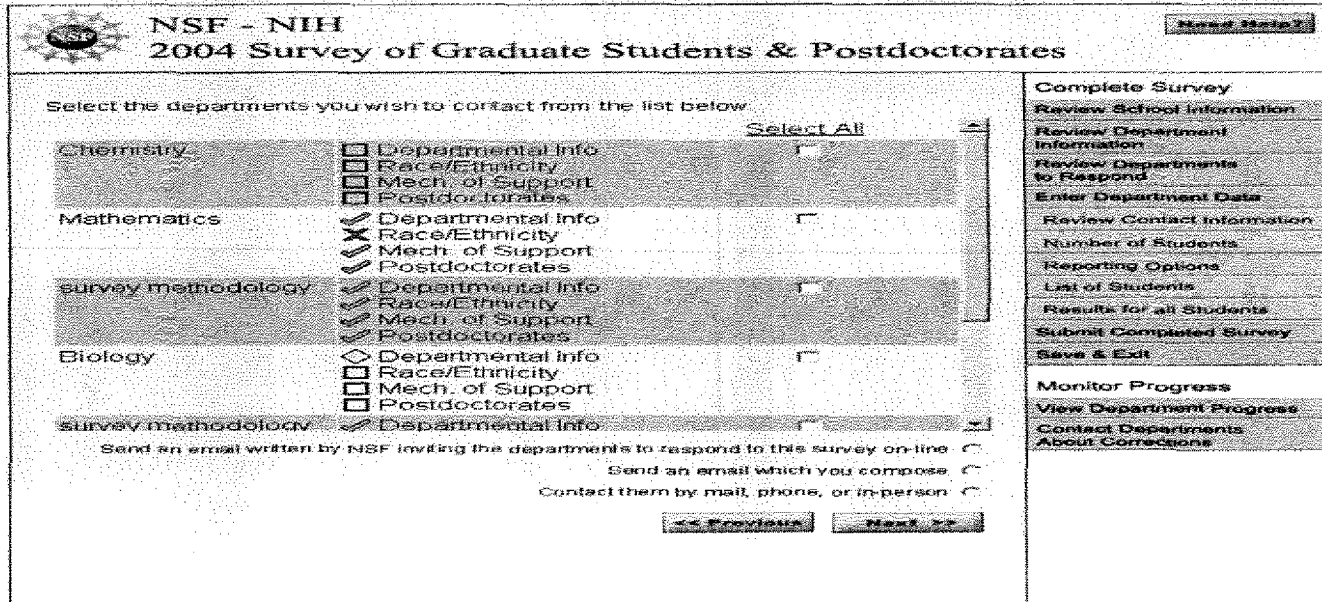


Figure 6. Unrestrained use of color proposed for Web version of the survey.

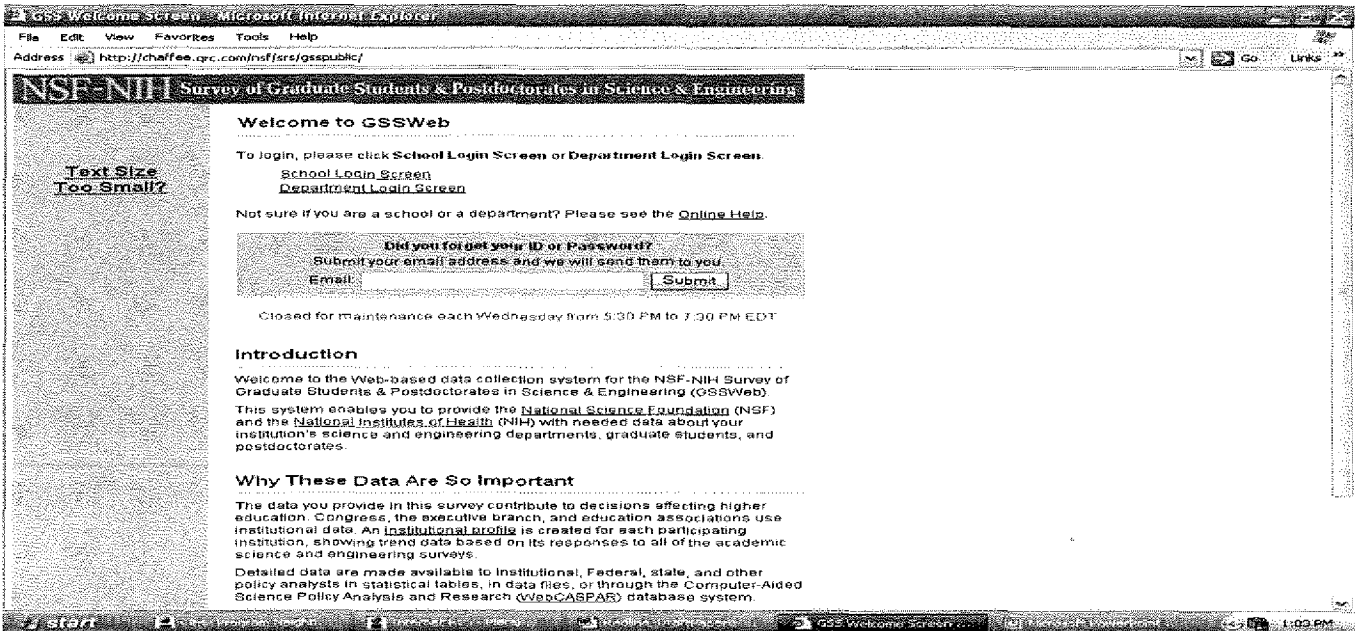


Figure 7. Original welcome screen.

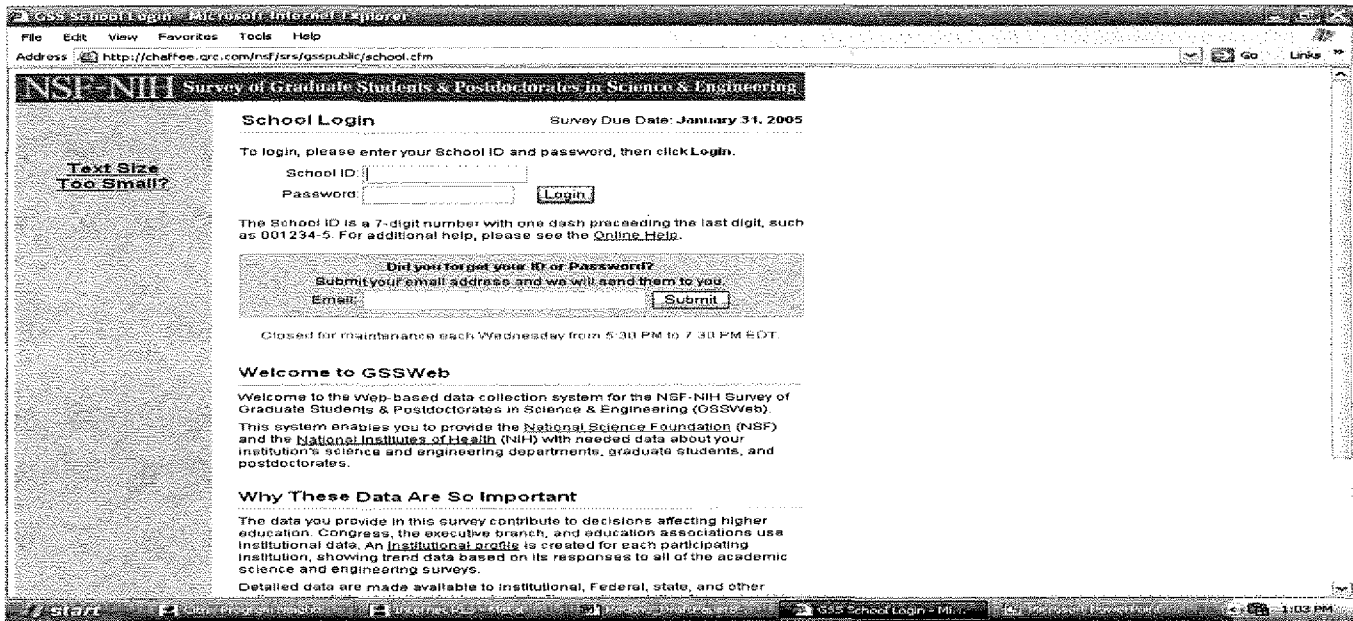


Figure 8. Original login screen.

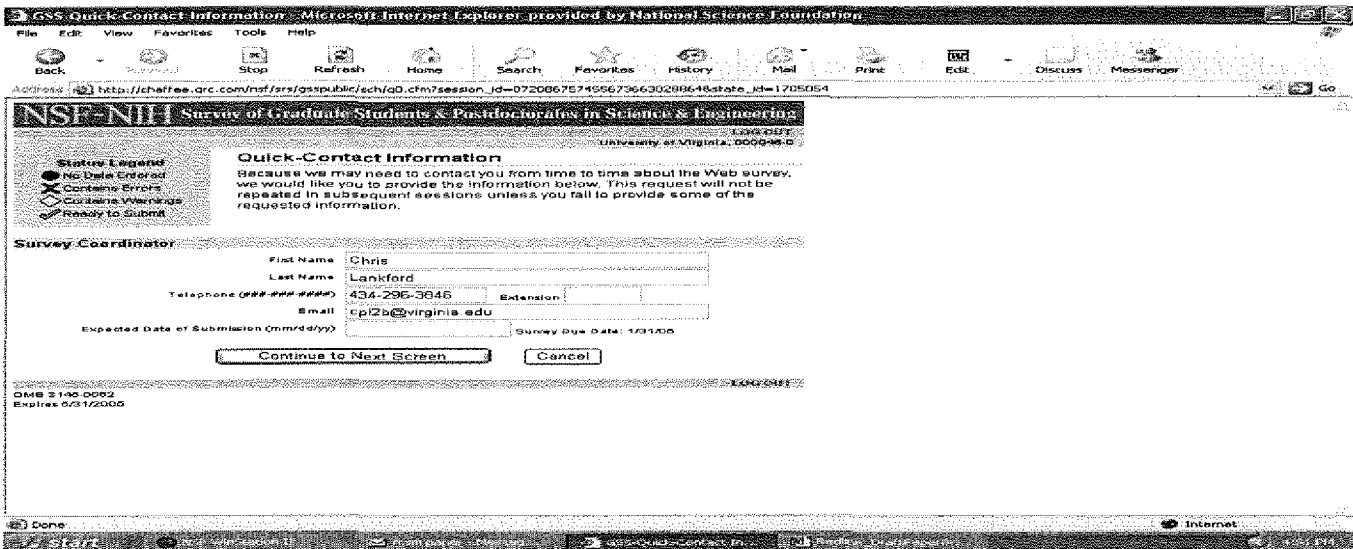


Figure 9. Original quick contact screen.

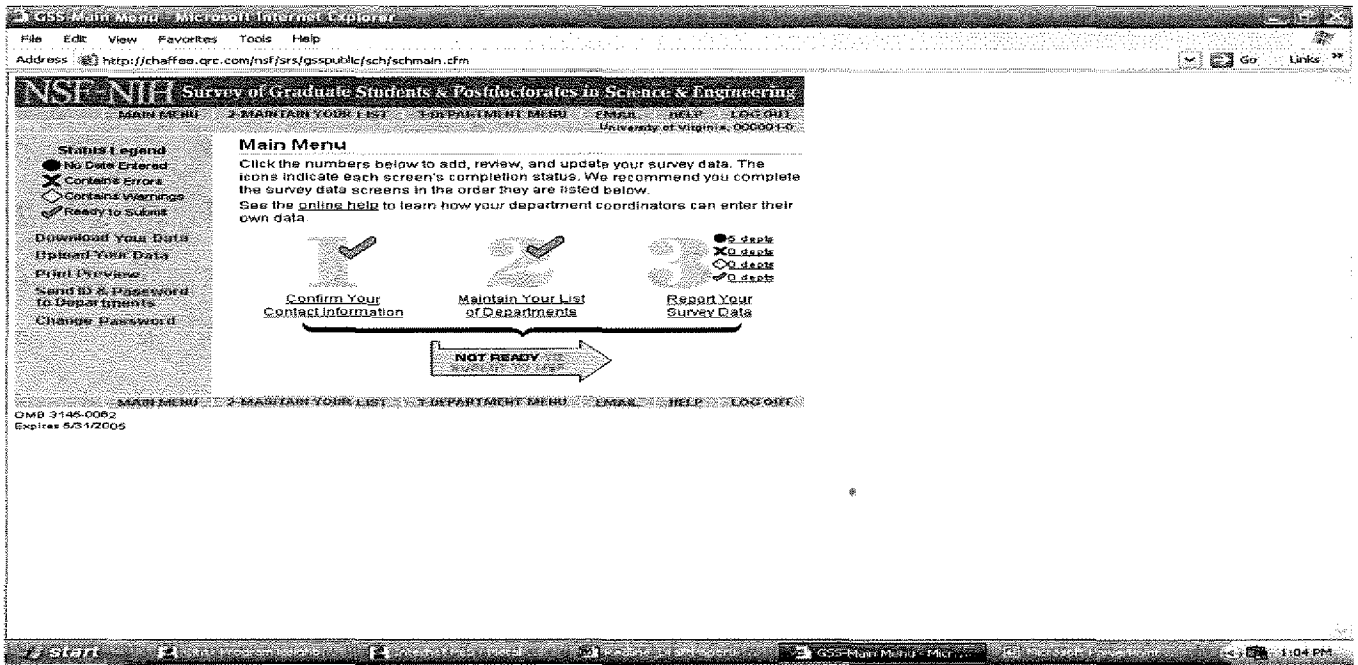


Figure 10. Original main menu screen.

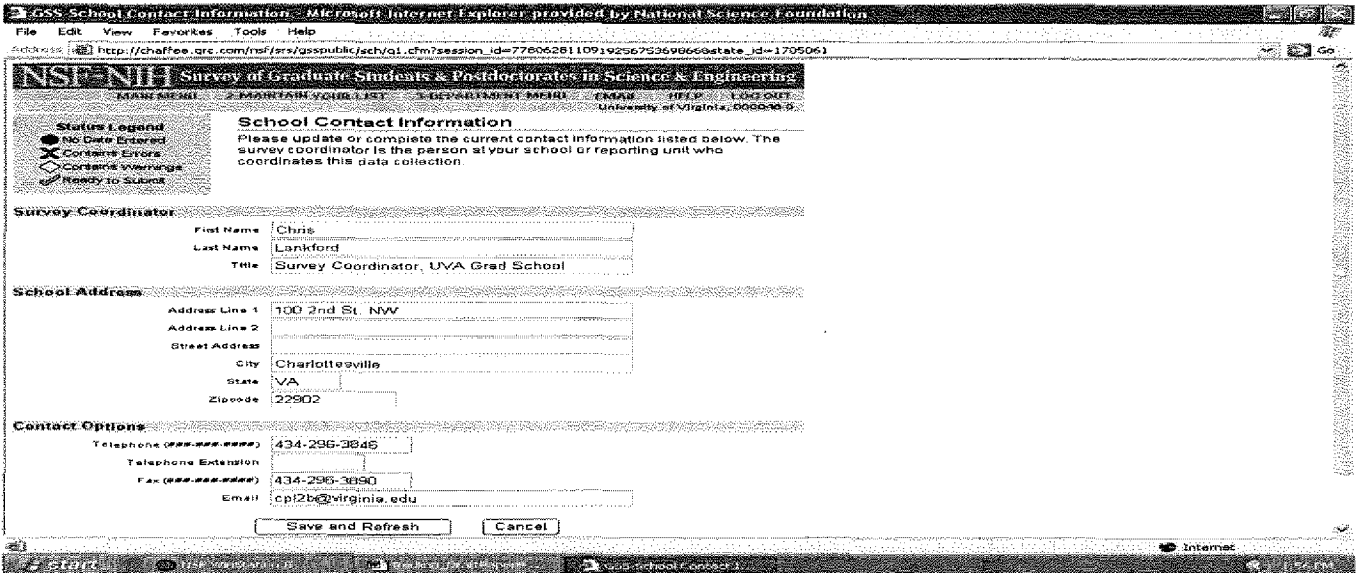


Figure 11. Original contact screen.

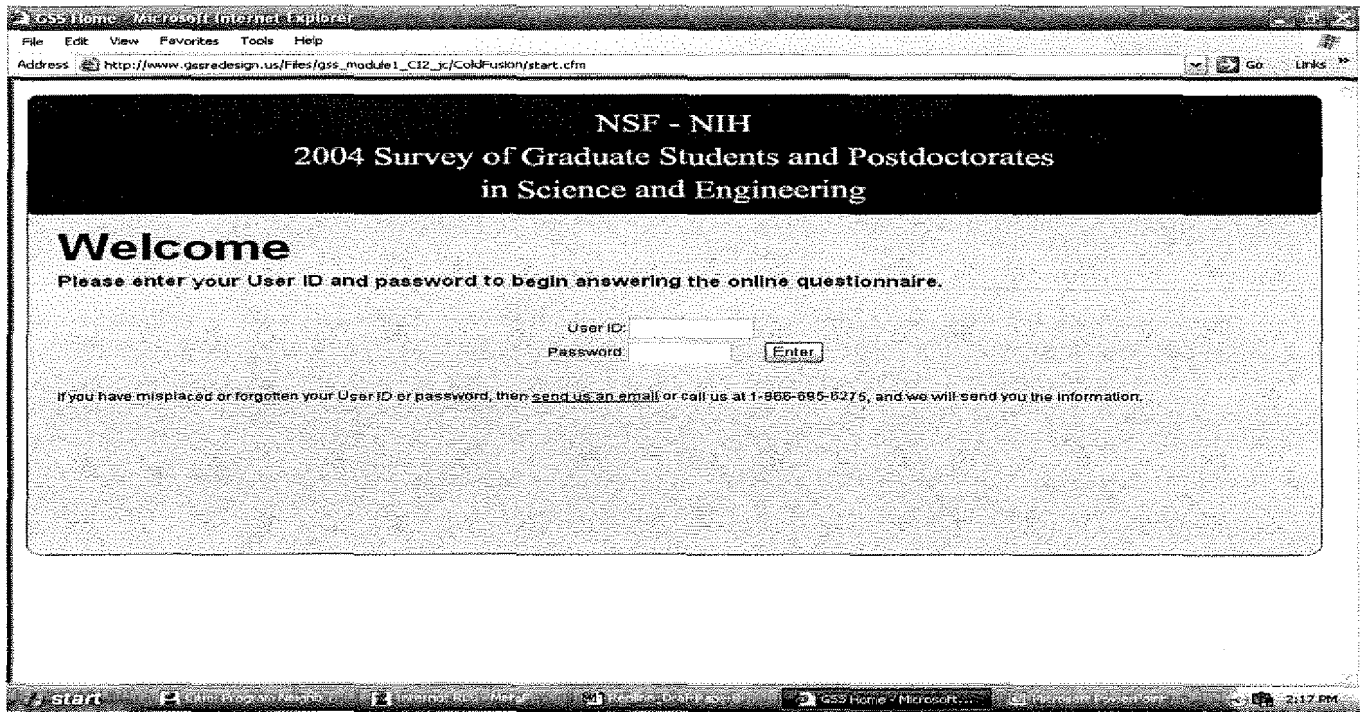


Figure 12. Redesigned welcome/login screen.

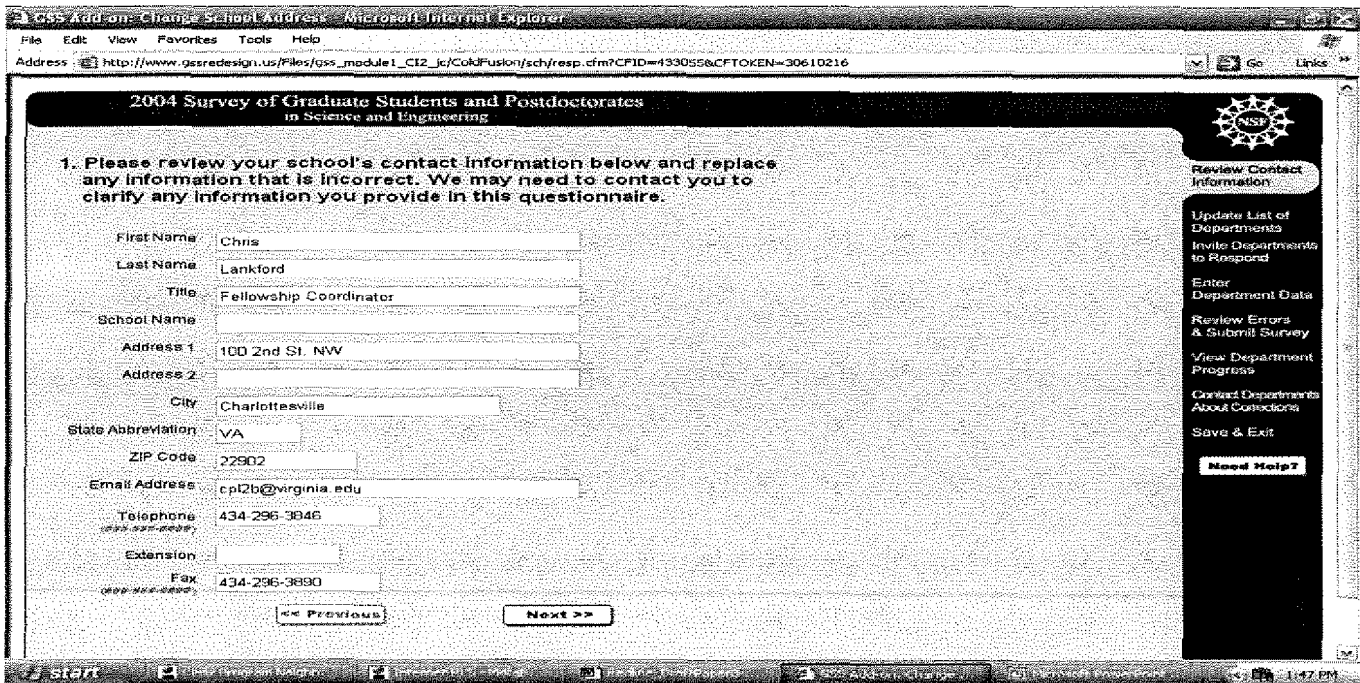


Figure 13. Redesigned contact item screen.

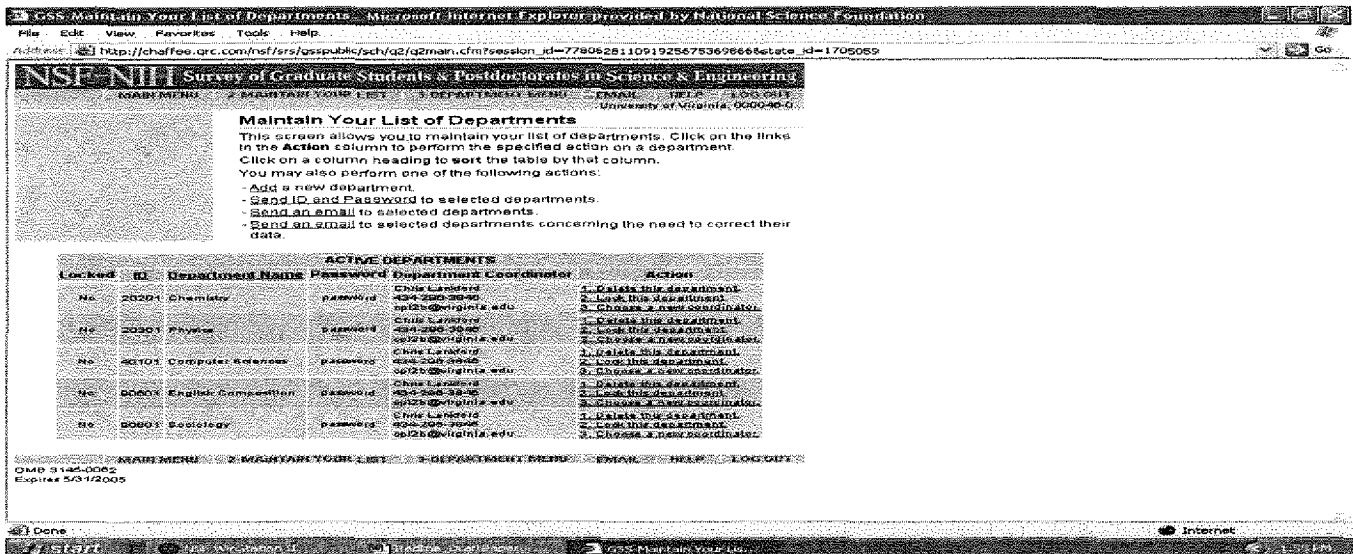


Figure 14. Original list of departments screen.

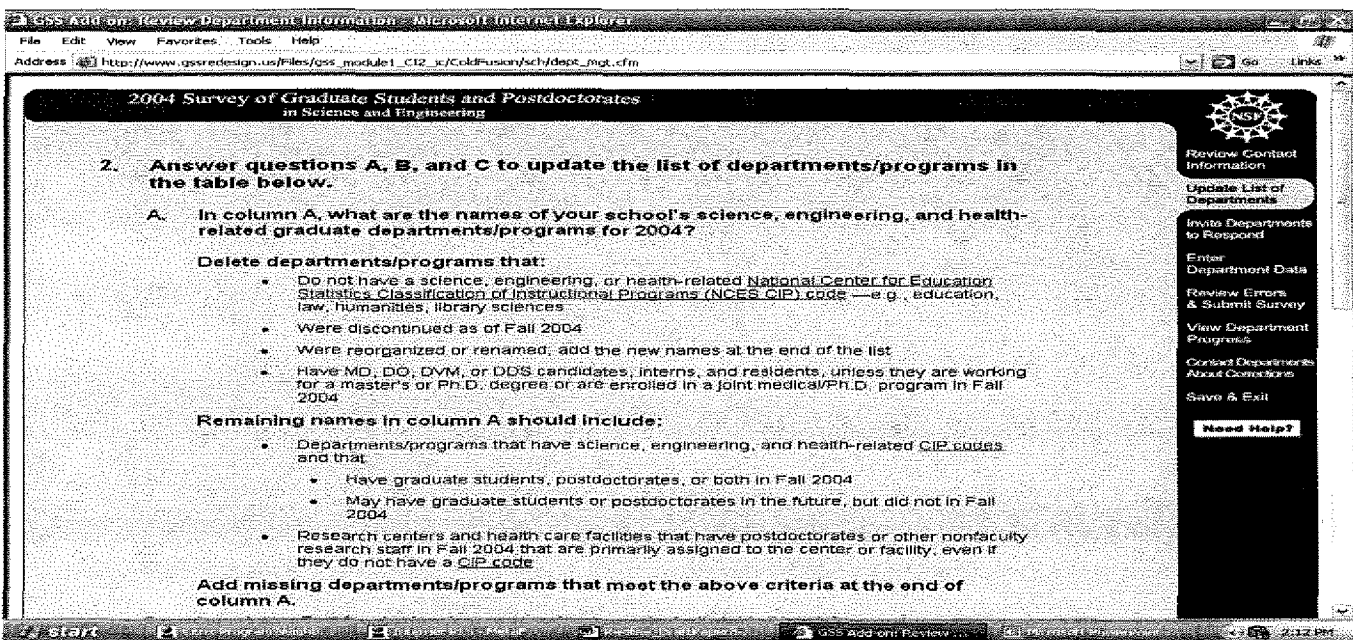


Figure 15. Top of the Redesigned List of Departments Screen.

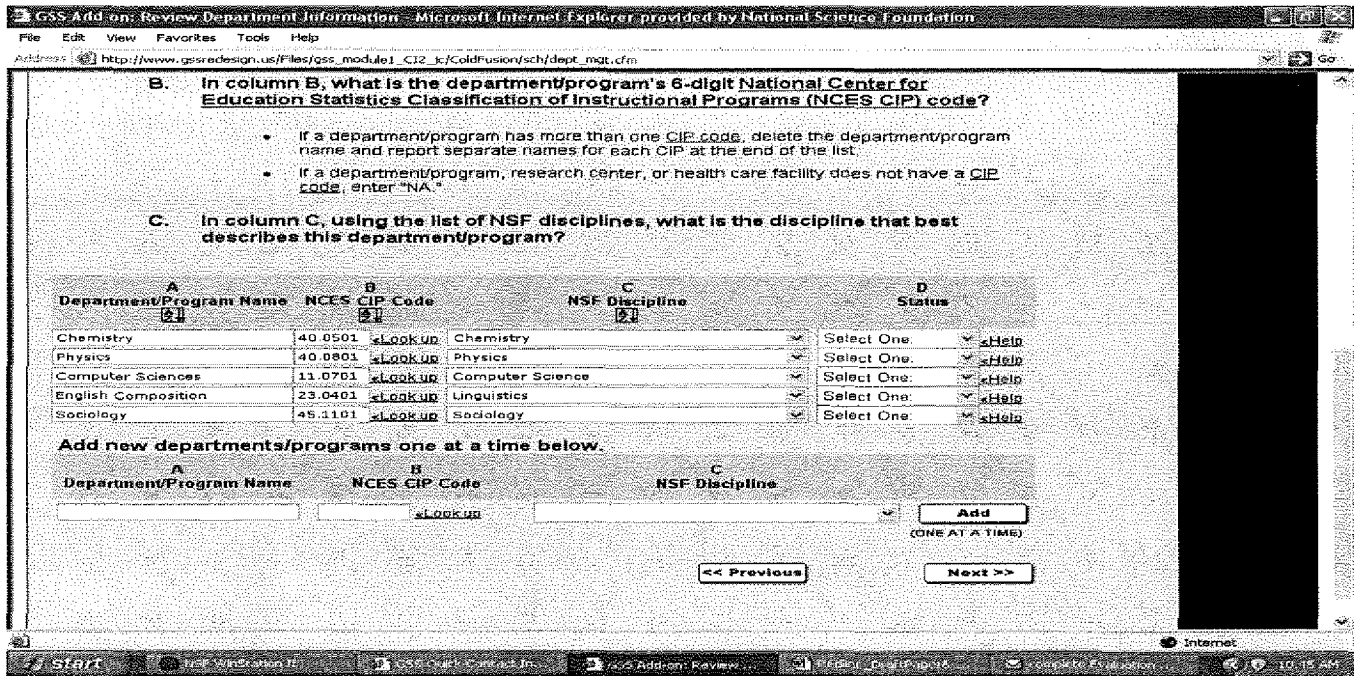


Figure 16. Bottom of the Redesigned List of Departments Screen.

Attachment 6.2

GSS Usability and Cognitive Interviews¹

The National Science Foundation (NSF) has contracted with RTI International to conduct research and redesign activities for the Survey of Graduate Students and Postdoctorates in Science and Engineering (GSS). One of the goals of the redesign effort is to improve data quality and minimize response burden. From December 2006 to May 2008, three rounds of usability/cognitive testing interviews were conducted on the GSS survey in an effort to identify areas of improvement in both the Form 811 and Form 812.

Rounds 1 and 2 of the GSS usability testing focused on modifications to the unit listing or Form 811. For Round 1, two variations of Form 811 in the 2006 GSS instrument were tested and evaluated. Following the first round of testing, a single data collection form was developed, incorporating the results of first round of testing. Round 2 testing was conducted on site at participating postsecondary institutions to ensure that the suggested revisions improved usability and reduced respondent burden.

Based on the results of the second round of usability testing, the data collection instrument was further revised and implemented as the 2007 GSS data collection instrument. The majority of the changes were made to the unit listing with only minor changes made to the data collection tables (Form 812). A third round of testing was then launched in the Spring of 2008 in an effort to evaluate respondents' reactions to the 2007 GSS survey and to gather insight for proposed changes to data collection tables (Form 812) that are being considered for the 2008 and future GSS surveys. *Table 1* provides an overview of the parts of the GSS instrument tested, the number of participants and the location of each round of usability/cognitive interviews. Each round of interviews is discussed in further detail in the following sections.

Table 1. Overview of usability/cognitive testing rounds

Round	Part of GSS Tested	Participants	Location(s)
1	Unit listing (Form 811)	12	Raleigh/Durham, NC
2	Unit listing (Form 811)	22	Washington, DC; Baltimore, MD; Philadelphia, PA; Los Angeles, CA; San Diego, CA
3	Unit listing (Form 811) & Data Collection tables (Form 812)	19	Minneapolis, MN; Atlanta, GA; Seattle, WA; Portland-Eugene-Corvallis, OR; and Pullman, Washington / Moscow, Idaho

¹ This is based substantially on a report prepared by RTI International for the NSF under contract SRS-0629305.

A. GSS Usability/Cognitive Interviews – Round 1

Participants and Methods

RTI methodologists recruited a total of twelve participants from universities within a reasonable driving distance from RTI that offered graduate programs in science, engineering, and/or health-related fields, including UNC-Greensboro, Duke University, UNC-Chapel Hill, North Carolina State University, and North Carolina Central University. Six of the twelve participants held positions at their institutions that qualified them as Institutional Coordinators (ICs), which means that they could provide data for all schools (e.g. graduate school, medical school). The other six participants were Sub-institutional Coordinators (Sub-ICs) and could only provide data at the school-level. The methodologists recruited both participants who had no prior experience with the GSS in any capacity, as well as those who currently serve, or served as GSS Coordinators. Of the twelve participants, seven had prior GSS experience. *Table 2* identifies test participants by institution, position at the institution, GSS role, and prior GSS experience listed by the order in which they were tested.

Table 2. GSS Round 1 Testing Participants

ID	Institution	School/Office	Role	GSS Experience
1	UNC Greensboro	Office of Institutional Research	IC	Yes
2	UNC Chapel Hill	School of Medicine	Sub-IC	Yes
3	UNC Chapel Hill	School of Public Health	Sub-IC	No
4	UNC Greensboro	School of Nursing	Sub-IC	Yes
5	UNC Chapel Hill	School of Medicine	IC	No
6	UNC Greensboro	College of Arts and Sciences	Sub-IC	No
7	NC State University	Office for University Planning and Analysis	IC	No
8	Duke University	School of Medicine	Sub-IC	Yes
9	NC State University	Graduate School	IC	Yes
10	Duke University	Graduate School	IC	Yes
11	NC Central University	Graduate School	Sub-IC	No
12	NC Central University	Research, Evaluation and Planning	IC	Yes

Usability testing for Round 1 was conducted at the RTI International main campus in Research Triangle Park, North Carolina in a private room with one-way mirrors and hard-wiring for audio and visual monitoring, enabling unobtrusive observation of interviews in the adjacent conference room. Upon arrival, each participant was greeted and escorted to the private conference room, where he or she met with the survey methodologist who conducted the usability tests. The survey methodologist read aloud the participant informed consent form, provided a copy to the participant, and asked for a signature and permission for observation.

The tests were conducted using RTI’s portable usability lab, a coordinated system of digital audio and video data capture equipment. The lab features professional-grade

video monitoring and recording capabilities, including a high resolution video camera. This equipment enables the methodologists to record each participant's on-screen activity, including button or link selection and keying, as well as viewing their facial expressions. The usability lab also allows the methodologist to code interviews in real time, recording errors, design and navigation issues, and questions, comments, and suggestions about the prototypes.

Usability testing focused on visual design, navigation, and ease of use of the survey forms. RTI developed usability protocols to collect information on participants' experiences using both the Base and Variant prototypes for the GSS. The protocols included scenarios that the participants were instructed to use to complete the survey. The scenarios simulated key tasks for Institution Coordinators and Sub-Institution Coordinators, including defining the organizational structure of an institution or sub-institution (a school/college/research center or institute) and assigning data reporting tasks.

In addition to examining usability, cognitive interviewing methods provided important tools for examining the thought processes that affect the quality of answers provided to survey questions and the completion of forms. Through the use of "think aloud" interviews, participants were able to tell the methodologists everything that they were thinking about while completing tasks for the Base and Variant. Concurrent and retrospective probes were used to gain insight on how respondents read, understood, navigated between, and entered data on the GSS screens. These probes were designed to provide information on issues with important implications for design, along with information on how participants with prior GSS experience have completed the GSS in the past. In addition, probes were developed to gain an understanding of the effects that different institution structures may have on GSS completion.

Results and Conclusions

Defining the Organization

The definition of "school/college" provided by the participants was consistent across institutions and matched the definition provided by the GSS. On the other hand, the definitions of department and programs of study varied significantly not only across institutions, but also within institutions (for example, Graduate School vs. IR Office). Some participants considered departments and programs of study to be equivalent, while others said that the difference between programs of study and departments was that departments offered degrees and programs of study did not. While some participants understood the nesting structure used in the GSS, they noted that many programs of study were either multi-departmental (interdisciplinary) or free-standing, complicating reporting under the Variant structure. Participants also noted that although an institution's organization fit into the GSS nested structure, many departments only have one program of study. One participant commented that while they use the nesting structure,

postdoctorates can only be associated with the departments and not the programs of study.

The ICs predominantly preferred the ability to upload their institutions' organization using the Variant prototype. However, they would prefer not to have to move through a number of steps or screens before they finished defining their institutions' organization. Overall participants thought the Variant prototype required them to go through too many screens and circle through the same screens multiple times. Conversely, participants thought the Base prototype required that too many tasks be completed on one screen, making it cumbersome to follow directions correctly. Therefore, we recommend finding a balance between the number of screens needed and the level of burden required on each screen.

With both surveys, participants had the most trouble understanding departments that had multiple programs of study or CIP codes associated with them. On the Base prototype, participants were inconsistent with how they identified departments with multiple CIP codes, and would sometimes only enter one CIP code for these departments. On the Variant prototype, participants were confused by the nesting structure that was assumed, especially when departments only had one program of study.

Making Task Assignments

When making task assignments, participants liked the flexibility offered on the Variant prototype, which allowed them to make some assignments at the institution level and other assignments at the school or department level. However, most participants also wanted to make all assignments at the same time rather than making institution assignments on one screen, school assignments on the next screen, and department assignments on yet another screen.

On the Variant prototype, participants did not understand the complete role of the coordinator. Only one participant recognized that a coordinator could either provide the actual data or make further assignments by departments or programs of study. All other participants thought that the coordinator's only role was to provide the data. Therefore, when participants reached the subsequent task assignment screens, they were confused because they thought they had already assigned the tasks. When explained, most understood the logic for breaking out assignments further for departments, but did not think it was necessary at the program of study level.

While participants appreciated the flexibility that the Variant offered for the test scenario, most indicated that for their institution, they would either provide all the data themselves or contact the departments to provide all the data. Only a few participants indicated that they might be able to provide some of the data (such as demographics) centrally for their institution, while contacting the departments for the rest (such as funding or postdoctorate information). Of the participants who would need to contact the

departments to get the data, a few indicated that they would email contacts outside of the survey to collect the data, but fill in the data themselves.

Making Modifications

For the most participants, updates and modifications to the organization or task assignments could be made easily using either survey. On the Base prototype, participants liked the use of the tabs on the right side of the screen. However, they thought that the tab labels, which were not specific enough, and should refer to individual tasks or screens rather than entire sections. On the Variant prototype, only half of the participants used the correct links to find the pages to modify their data. Once directed to the correct screens, they could easily update the organizational structure or task assignments. In addition, participants liked the home page on the Variant prototype, but could not always get there easily.

Finally, participants wanted the ability to explicitly save their data before exiting the survey. While the Variant prototype automatically saves the participant data, users were not aware of this and were concerned about exiting without actually hitting a save button.

Comparisons by participants

The twelve participants who completed usability testing were from five different universities, a variety of departments and offices within their universities, with differing levels of knowledge about the GSS. Despite these differences, participants generally responded to most of the tasks in a similar manner. The following sections describe the few differences observed during testing by the type of participant.

Participants with GSS Experience versus Participants without Experience

Of the twelve participants who completed the usability tests, seven had previous experience as GSS coordinators. We anticipated that participants with prior GSS experience would be more familiar with the demands of the GSS, but this was not entirely true. Of the four participants who were not familiar with CIP codes, three had previous GSS experience. Similarly, participants without GSS experience were as familiar with their institutions' organization as participants with experience. In fact, two participants with GSS experience indicated that they do not update the list of departments each year, but merely use the same list as they had used the year before. They commented that if they had to start from scratch, they would not know which departments to include or exclude for the GSS. Specifically, one GSS coordinator indicated that she only reports data on 12 departments, and that these are the same 12 departments for which they have reported data on over the past 20 years.

Participants without GSS experience tended to read the instructions more and make fewer assumptions than participants with GSS experience.

Institutional Staff versus School/College Staff

Of the twelve participants, three were from an office of Institutional Research (IR) or equivalent, three were from a graduate school, three were from a medical school, one was from a nursing school, one was from a College of Arts and Sciences and the other was from a School of Public Health. In general, staff from IR offices were more knowledgeable about the institution's organizational structure than staff from the graduate schools or other schools or colleges. All IR staff were familiar with CIP codes, while one third of the school-level staff were not.

However when it came to task assignment both IR staff and school/college level staff primarily got the data from the departments, either by emailing the departments through the survey or emailing them outside of the survey.

ICs versus Sub-ICs

Sub-ICs did not see the "Organize your Institution" screen, and therefore, did not receive the instructions on how the GSS is organized. Sub-ICs who were not familiar with the department – program of study structure assumed by the GSS had a more difficult time using the screens that called for programs of study to be defined for each department.

In general, Sub-ICs, particularly those with GSS experience, were more likely than ICs to pass off control of the GSS by assigning data collection to the departments. Five of the six sub-ICs said that they would email the departments and have them fill out the GSS. The other sub-IC said that he would have the departments send him a copy of their reports and he would fill it out himself. None of the sub-ICs said that they could provide any of the data centrally for their school or college without contacting the individual departments. Two sub-ICs with prior GSS experience said that their role on the GSS was strictly to email departments and make sure they completed the survey, but they did not know anything about how the departments completed the GSS. ICs, on the other hand, preferred to have more control over the GSS, either by indicating that they could provide some of the data centrally themselves or that they would email departments, off-line, and enter the data themselves.

B. GSS Usability/Cognitive Interviews – Round 2

Following the first round of testing, the data collection form revised and tested with a new set of participants in their offices to better replicate the environment for the GSS.

Participants and Methods

For the second round of usability tests, a total of 22 interviews were conducted

on-site at 16 institutions across the country. Of the 22 tests, 18 were conducted with the electronic version of the instrument. An additional 4 tests were conducted with the hardcopy version of the Form 811. However, one participant completed both an electronic and a hardcopy usability test. Therefore 22 tests were conducted with 21 participants. **Table 3** identifies the 18 Web participants by institution, coordinator type, title, and GSS experience. Some of the participants indicated that while they met the criteria for IC or SC, the survey would actually be completed by their administrative assistant or support staff. In those instances, we conducted the usability test with the person who would actually complete the GSS surveys, while collecting qualitative information from the recruited coordinators, when available. For these cases, the title of the recruited coordinator is listed first in the table, followed by the title of the person who actually completed the usability test.

Table 3. Round 2 Web Participants

ID	Institution	Coordinator type	Title/Office	GSS Exp. (Years)
1	Howard	School of Med	Office of the Dean of Research	10+
2	Howard	Institution	Office of the Dean for Educational and Research Affairs	0
3	Johns Hopkins	School of Med	Office of the Dean for the Registrar/Faculty Records Administrator	12/1
4	Johns Hopkins	Institution	Provost Office	3
5	Morgan State	Institution	Office of Institutional Research	0
6	Thomas Jefferson	Institution	Graduate School Dean's Office	4
7	Rutgers, Camden	Institution	Graduate Dean's Office	2/1
8	Drexel	Institution	Office of Institutional Research	0
9	Penn	Institution	Provosts Office	0
10	Temple	School of Med	Office of the Dean of Graduate Studies, School of Medicine	2
11	Temple	Institution	Office of Institutional Research	4+/4
12	Loyola Marymount*	Institution	Officer of Institutional Research	0
13	USC	Institution	Program Manager	6
14	UC San Diego	School of Med	Office of the Dean, School of Medicine	10+/1
15	Scripps	Institution	Office of the Dean of the Graduate Program	0

16	UC San Diego	Institution	Office of Graduate Studies	1
17	National	Institution	Office of Institutional Research	1
18	Univ. of San Diego	Institution	Office of Institutional Research and Planning	2

Table 4 identifies the four hardcopy participants by institution, coordinator position, title, and GSS experience.

Table 4. Round 2 Testing Hardcopy Participants

ID	Institution	Coordinator type	Title/Office	GSS Exp. (years)
19	Temple	School of Pharm.	Office of the Dean, School of Pharmacy	4
12	Loyola Marymount*	Institution	Office of Institutional Research	0
20	UC Irvine	School of Med.	Office of the Dean, School of Medicine	11
21	San Diego State	School of Eng.	Office of the Dean, School of Engineering	5+

*Participant also completed a usability test for the Web survey.

Prior to beginning the usability test, all participants were asked a series of pretest questions about their experience with the GSS, their knowledge of the GSS, what GSS data they could provide, and how their institution was organized with respect to graduate students, postdoctoral appointees, and nonfaculty researchers. Following completion of the pretest questionnaires, participants were then asked to complete the redesigned Form 811 either on the web or on hardcopy. Survey methodologists observed participants interacting with the instrument, noting how well they were able to navigate the survey, which features they used, errors they made, and questions they had. The survey methodologists also employed common cognitive interviewing strategies such as asking participants to think aloud while they completed the survey as well as probing the participants about some of the questions on the instrument.

In addition to the 22 usability tests, 3 people were interviewed about how data on postdoctoral appointees were collected at their institution. The interviewees were selected because they were identified by the NSF or usability test participants as knowledgeable about postdoctoral appointees

Results and Conclusions

The redesigned Form 811 consisted of a number of significant changes from the 2006 GSS survey that were tested and evaluated. One change was the introduction of a new coordinator role, called the Institution Coordinator (IC), for institutions containing

multiple schools/centers (e.g., institutions containing a medical school and a dental school in addition to the graduate school). The IC role was added to increase the awareness of the GSS among institution-level personnel, and so that one person would be responsible for coordinating all aspects of GSS data collection for the institution – starting with the listing of all schools and research centers that are relevant to the GSS. The aim is to increase awareness of the GSS and to maximize the likelihood of full coverage of GSS-relevant schools, research centers, departments, programs, etc. The methodologists recruited participants from the Institutional Research Office, Dean’s Office, or Provost’s Office to serve as IC. However, several of the participants who actually completed the usability test were not the original people recruited, but an administrative assistant or support staff person who was presumably not as knowledgeable about the institution. Most of the remaining ICs had been in their position at the institution for one year or less. As a result, the participants who served as ICs during testing were not very familiar with the GSS or their institution’s organization. Of the 14 ICs, 8 were not familiar with the GSS, 9 were not familiar with CIP codes, and 9 indicated that they did not readily know their organization’s complete structure, but could probably find out. Because usability testing was only done with a small convenience sample of institutions, it is unclear at least from this convenience sample, if it would be feasible to identify one person at most institutions who would be knowledgeable enough to serve as Institution Coordinator in the capacity intended.

The introduction of the IC corresponds with a four-tiered data collection structure that was also introduced in the redesigned Form 811: institution-level; school-/center-level; department-level; and field-of-study-/program-level. The 2006 GSS data collection has a two-tiered structure: schools and departments. GSS school coordinators (SCs) often work in a school-level office such as a graduate school or a medical school. The SC updates the list of departments in their school each year and may provide the data themselves or assign personnel within departments to do so. Departments are the second data collection tier for the current GSS. The four-tier data collection structure added an institution level above schools and a field of study level beneath departments. The four-tiered organizational structure proved difficult for some participants during testing. While most participants were familiar with the terms “school” and “department,” they used varying terms for “field of study” such as “program,” “division,” “track,” or “concentration.” The four-tiered structure introduced challenges for some institutions because they did not have any “fields of study.” It also became clear that most institutions had many “exceptions to the rule.” That is, even when institutions had fields of study, not all fields of study were treated the same. Some fields of study were considered independent of departments, some were tied to multiple departments, some were within departments, some departments only had one field of study, and some had multiple fields of study. The redesigned Form 811 did not take into account all of the variations to the four-tiered structure. As a result, some participants had a difficult time using the web survey to enter departments and fields of study. Despite the differences in terminology and the many exceptions to the rule, participants indicated that if better instructions and

definitions of the terms were used, they could map their institution to the four-tiered structure of the redesigned Form 811.

In addition to asking for a four-tiered structure, the redesigned Form 811 also asked ICs to identify CIP codes for all fields of study. Because only 5 of the 14 ICs were familiar with CIP, this was difficult for many respondents. In addition, the web survey screens asking about CIP codes were not easy for participants to navigate, making the task harder. Participants looked through the list of codes and made their best guess as to what fit. It was unclear to participants how to proceed if there was not a CIP code that matched. Rather than deleting that field of study, participants tended to assign the closest CIP code that matched. Using the manual web screens to identify CIP codes was a tedious and burdensome task for participants. Again, many of the participants in this usability test were administrative assistants or new to their position. It is possible that another person at the institution would be more knowledgeable about CIP codes. Nonetheless, the burden associated with manually entering the entire set of departments and programs for a sizeable institution would be great in any case. Alternative approaches should be considered and further investigation is warranted.

Another new feature of the redesigned Form 811 is that it allowed ICs to upload their organizational structure electronically. It was hoped that this would alleviate the burden of having to manually enter all departments and fields of study into the web program. Due to time restraints, only three of the 14 ICs were able to provide an uploaded file. However, seven other participants commented that uploading would be a desirable option. Only one participant stated that manual entry was preferred due to the relatively few graduate programs offered at the institution. Two of the participants who provided an electronic file were familiar with CIP codes and used a centralized database to develop the file. They indicated that providing a file was easy. The other participant created the file manually, which was more labor intensive. Only one of the electronic files was tested with school coordinators for that institution. The school coordinators identified mistakes in the uploaded file. More research will need to be conducted to determine what percentage of institutions can *easily* provide an electronic file and to determine how accurate the files are. Alternative approaches (e.g., utilizing existing GSS departments and linking those departments to Integrated Postsecondary Education Data System [IPEDS] Completions programs) should be investigated.

Another feature of the redesigned Form 811 was that it allowed for more flexibility in assigning data collection tasks. With the current GSS, the school coordinator can either provide all of the data (i.e. demographic and financial information about graduate students, postdocs and nonfaculty researchers) or the coordinator can assign a department respondent to provide it. The redesigned Form 811 allows the Institution- or School- Coordinator to identify what pieces of data they can provide at the institution- or school- level (e.g., demographics) and then assign the remainder of the data collection tasks to a department coordinator/respondent and/or other institution personnel. Ten of

the 21 participants indicated this level of flexibility was unnecessary as they merely emailed the departments and did not provide any data themselves. The flexibility was beneficial to five participants, who felt they could provide certain pieces of the data themselves – usually graduate student demographics and sometimes postdoctoral appointee demographics or graduate student financial information. The remaining 6 participants indicated that they could either provide all of the data themselves or would enter it themselves even if the departments actually produced the information. Overall, participants reacted positively toward the added flexibility, but thought that the task assignment section had too many screens and that the text was sometimes confusing. Alternatives to simplify this tool while allowing for the increased flexibility, such as changing the “task assignment” activity to an only-as-needed “delegation of responsibility” activity, should be investigated.

Participants also commented that they wanted to know at the beginning of the survey what was coming next. The redesigned Form 811 was quite different from previous versions of the GSS, and it was not always clear how it was different. To address this, participants would like to download and print a hardcopy version of the survey. Furthermore, participants were confused about what departments and fields of study the GSS was interested in. Participants need to be provided with clearer guidelines about what should be included and what should be excluded.

Results from this study are not intended to generalize to a larger population, but rather reflect how typical participants use and respond to the redesigned Form 811, and to identify areas for further investigation.

Changes to the 2007 GSS

As a result of information gathered during the usability and cognitive testing interviews that were conducted from December 2006 to May 2007, RTI revised the 2007 GSS survey. The 2007 GSS survey differed from the 2006 GSS survey in several important ways. The site visits revealed that participants did not fully understand the purpose of the unit listing and were not updating it correctly. Therefore the 2007 GSS emphasize the unit listing (Form 811) by dividing the 2007 GSS survey into two separate parts with two deadlines. The unit listing or Form 811 was Part 1 of the survey. The data collection tables or Form 812 were Part 2 of the survey. In addition, more detailed instructions were provided for the unit listing and an expanded set of crosswalks for identifying GSS-eligible units was added.

We also learned from the site visits that the terminology used by institutions varies considerably. A “department” at one university might mean something entirely different at another university. Therefore a change in terminology was made from departments/programs to a more comprehensive term “organizational units”. A revised method for assigning unit respondents was implemented in order for school coordinators to quickly assign respondents and to ensure that respondents received the notification. Finally, the survey deadline was extended by one month to assist unit respondents in submitting their data by the deadline. In addition to these

major structural changes, the 2007 survey also had a number of minor design changes such as the use of a home page, menu bars, expanded glossary, print functions, calculate totals buttons, and other features.

Part 1 of the 2007 GSS web survey was completed by the School Coordinators (SCs). The SCs completed a series of tasks to initiate the data collection at their school. They updated their contact information (Coordinator Contact Information screen), updated their list of departments, programs, research centers and health-care facilities for 2007 (Update Your Unit List screen), updated unit respondent assignments (Update Unit Respondents screen), were reminded to inform unit respondents of their GSS assignment, (Notify Unit Respondents screen), prompted NSF to email Part 2 IDs and passwords to unit respondents (Send or Resend ID/Password) and submit Part 1 by the deadline of November 30 (Submit Part 1 screen). The interim deadline was added in the 2007 data collection to encourage school coordinators to notify unit respondents about the survey in a timely fashion.

Part 2 of the 2007 GSS web survey included the data collection tables for the individual units and was completed by either Unit Respondents (URs) or School Coordinators who also completed the unit-level information (SC/URs). The three data tables were Race/Ethnicity and Citizenship, Sources and Mechanisms of Support, and Postdocs and Nonfaculty Researchers. In addition to the three data tables, Part 2 also included a screen to provide contact information and the unit's highest degree granted (Unit Profile screen). Unit respondents were able to monitor their progress and errors for their assigned unit(s) and submit completed error-free units on the View and Submit Data screen. On this screen, SCs were able to monitor the progress of all units on their unit list and submit complete and error-free data to NSF. The deadline for submitting Part 2 was February 29, 2008 which was a month later than previous years.

School coordinators also had access to two additional screens in Part 2 that unit respondents did not. The Upload Data screen allowed SCs to upload their data instead of keying in their responses. The Download Copy of Survey Data allowed SCs to obtain a comma-delimited file of their data for the current and previous year (2007 and 2006).

C. GSS Usability/Cognitive Interviews – Round 3

In 2007, RTI conducted site visits, usability tests and cognitive interviews with colleges and universities across the country to gather information on how respondents complete the GSS survey and how it could be improved. As a result, the GSS survey was redesigned in 2007 to improve data quality and minimize response burden. The majority of the redesign activities were focused on improving the unit listing and updating activities (Form 811). However, small changes were also made to the data collection tables (Form 812). This report focuses on a round of usability tests and cognitive interviews conducted on the 2007 GSS survey in an effort to evaluate respondents' reactions to the revised survey and to gather insight for proposed changes that are being

considered for the 2008 and future GSS surveys. Highlights of our methods and results are included in this summary; details follow in the rest of the report.

Usability and cognitive testing for this project consisted of two parts. The first component of the testing was to evaluate users' reaction to the 2007 GSS Web Survey. Respondents were asked to recomplete the web portion of the survey that they completed in 2007. This allowed methodologists to evaluate how the changes to the GSS website affected how respondents completed the 2007 survey. This part of the testing will be referred to as "2007 GSS Web Survey" usability/cognitive interviews.

The second part of the testing was to evaluate proposed changes to the data collection tables (Form 812) to ensure that the revisions improved data quality without increasing respondent burden. The proposed changes for the 2008 GSS Survey (and future years) were presented to respondents on the hardcopy worksheet. The proposed 2008 changes applied only to Form 812, therefore the worksheet was not tested with respondents who completed the unit listing only. This part of the testing will be referred to as the "2008 GSS proposed changes" cognitive interviews. Both components of testing occurred on site at 14 different institutions in five different regions across the country: Minneapolis, MN; Atlanta, GA; Seattle, WA; Portland-Eugene-Corvallis, OR; and Pullman, Washington/Moscow, Idaho.

2008 GSS Proposed Changes

The hardcopy worksheet corresponds with Part 2 of the web survey. The worksheet is intended to be used by unit respondents as needed to prepare their responses before entering their data into the tables in Part 2 of the web survey. Therefore, any revisions that are made to the data tables in the worksheet for the 2008 data collection will also be made to the corresponding data tables in the web survey. Based on feedback received from two focus groups with data users and reviews by survey methodologists, the 2007 worksheet was revised and subject to cognitive testing with unit respondents. The modifications to the worksheet included adding rows and reorganizing columns in tables; revisions to definitions; and rewording of questions, instructions, rows stubs and column headings. Some of the more significant changes are described here.

The three data tables were modified in an effort to simplify the reporting task for unit respondents. The first table (Question 2) collects counts of graduate students by race and ethnicity, citizenship, enrollment status and sex. The race and ethnicity columns were reorganized in an effort to clarify the distinction between race and ethnicity and to align with the categories institutions will be required to use for the Integrated Postsecondary Education Data System beginning in 2010. In 2007 and prior years, the table was organized such that graduate students (who were U.S. citizens or permanent residents) were first divided on the basis of having a single race or more than one race and whether they were non-Hispanic or Hispanic. If they were non-Hispanic and one race, they were further divided by racial category. The revised scheme first categorized

graduate students as non-Hispanic or Hispanic. If they were non-Hispanic, they were categorized by racial category or as multiracial. Hispanic students were to be reported together whether they were of a single race or multiracial. The instructions for question 2 were revised accordingly. In addition, the glossary added an entry for “Hispanic or Latino” and “Ethnicity.”

The second table (Question 3) gathers information on graduate students’ sources and mechanisms of financial support. It was modified such that “self-support” is a third category on the level of federal sources and nonfederal rather than a category within nonfederal sources. Also, the cell corresponding to HHS/NIH graduate teaching assistantships was blackened so that erroneous data would not be entered; HHS/NIH does not fund graduate teaching assistantships.

All three revised tables added a row for counts of men. Previously, all tables provided a row for totals (men and women) and directly beneath a row for women. While the number of men could be calculated by subtracting the number of women from the total, the number of men was not explicitly called for in the table. This created confusing error messages in the web survey (e.g., The value of full-time men (Row 4) must be greater than or equal to the calculated value of first-time men (Row 7 minus Row 8)) and required special programming code for those who collated the data electronically. Therefore, a row was provided to enter counts of men to each of these tables.

Several definitions and terms were revised for clarification. The definition of first-time graduate students used in previous years of the GSS was ambiguous in that it was unclear whether “first-time” was modifying the pursuit of a graduate degree or enrollment at the institution. The definition and corresponding instructions were revised in an effort to clarify that the intent was to identify students who were enrolled at the GSS institution for the first time, even if they had enrolled at another institution previously. The revised worksheet also redefined students doing their research away from campus. Previously, students doing thesis or dissertation research in a foreign country were not intended to be counted. The instructions were revised such that these graduate students would be included. Also, throughout the worksheet the term “foreign” was replaced by “foreign nationals holding temporary visas.” Finally, the definition of the length of postdoc appointments was changed from “generally 5-7 years” to “generally no more than 5-7 years.”

Participants and Methods

The usability/cognitive testing for the 2007 GSS Web Survey was designed to replicate a user’s real experience completing the survey. As a result, the interviews were conducted on site in the participants’ office or a small conference room, allowing participants access to data and resources that might be necessary to complete the survey. The tests were conducted at 14 different institutions in five different regions across the country: Minneapolis, MN; Atlanta, GA; Seattle, WA; Portland-Eugene-Corvallis, OR; and Pullman, Washington/Moscow, Idaho.

Before beginning the interview, the survey methodologists who conducted the interviews read aloud the participant informed consent form (Appendix A), provided a copy to the participants, and asked for a signature and permission for recording. Participants in the 2007 GSS Web Survey interviews were asked to log on to a test version of the survey using their own computer. The test version of the site replicated the survey that they saw at the beginning of data collection rather than the completed survey. Participants were asked to take their time to complete the survey, providing real data as if they were actually completing the GSS. While the participants completed the Web survey, one methodologist observed and recorded the participant's on-screen activity, including button or link selection and keying to code interviews in real time for reporting errors, design and navigation issues, questions, comments, and suggestions about the survey.

In addition to examining the usability of the site, the methodologists employed cognitive interviewing methods to examine the thought processes that affected the quality of answers provided to survey questions and the completion of forms. Through the use of "think aloud" interviews, participants were encouraged to tell the methodologists everything that they were thinking about while completing tasks. Concurrent and retrospective probes were used, as necessary, to gain insight on how respondents read, understood, navigated between, and entered data on the GSS screens. These probes were designed to provide information on issues with important implications for design, along with information on how participants with prior GSS experience have completed the GSS in the past. In addition, probes were developed to gain an understanding of the effects that different institution structures may have on GSS completion. However, the probes were used sparingly so that they would not disrupt the survey process.

To test the 2008 GSS proposed changes, participants were shown a revised hardcopy worksheet. Participants were asked to read through the worksheet, and answer questions aimed to elicit whether participants understood the proposed changes and what the effects would be on them. During this part of the interview, participants were not asked to provide any real data, just to provide their reactions to the proposed survey changes.

Usability Participants

RTI Methodologists recruited participants from 14 different universities and colleges (institutions) from five different regions across the country: Minneapolis, MN; Atlanta, GA; Seattle, WA; Portland-Eugene-Corvallis, OR; and Pullman, Washington/Moscow, Idaho. **Table 5** shows the selected institutions by the following characteristics:

- Rank by graduate student total enrollment;
- Institutional control (public vs. private);
- Carnegie Classification (e.g., Research Universities – very high research activity; Special Focus Institutions—Medical schools and medical centers);
- Historically black colleges and universities (HBCU).

Table 5. Participating Institutions

University	City	Control	Enrollment Rank	Total student enrollment	Carnegie	HBCU
Emory University	Atlanta, GA	Private	150	10,000-19,999	Research University	No
Georgia Inst. of Technology	Atlanta, GA	Public	12	10,000-19,999	Research University	No
Georgia State University	Atlanta, GA	Public	113	20,000 & above	Research University	No
Mayo Graduate School	Rochester, MN	Private	323	Under 1,000	Doctoral/Research Univ.	No
Morehouse College	Atlanta, GA	Private	367	Under 1,000	Medical school/center	Yes
Oregon Graduate Institute	Portland, OR	Public	339	1,000-4,999	Medical school/center	No
Oregon State University	Corvallis, OR	Public	82	10,000 - 19,999	Research University	No
Portland State University	Portland, OR	Public	102	20,000 & above	Doctoral/Research University	No
University of Idaho	Moscow, ID	Public	176	10,000-19,999	Research University	No
University of Minnesota	Minneapolis, MN	Public	3	20,000 & above	Research University	No
University of Oregon	Eugene, OR	Public	223	20,000 & above	Research University	No
University of Washington	Seattle, WA	Public	9	20,000 & above	Research University	No
Walden University	Minneapolis, MN	Private	1	20,000 & above	Doctoral/Research University	No
Washington State University	Pullman, WA	Public	114	20,000 & above	Research University	No

Usability Test Sample

As described above, participants were identified as either school coordinators (SCs), unit respondents (URs) or school coordinators who also served as unit respondents (SC/URs) depending on their role at the university and in completing the survey. **Table 6** provides a description of the usability participants by coordinator type and part of the interview completed. During the recruitment process, we identified four school SCs, seven URs, and eight SC-URs. However, during testing it was revealed that two of the SCs entered the data for all units into the GSS survey, but the data was actually provided by department-level contacts.

Table 6. Usability Test Participants by Mode and Coordinator Type

Interview Section	Number of SCs	Number of URs	Number of SC/URs	Total Number
2007 GSS Web Survey	4	6	8	18
2008 GSS Proposed Changes	2	7	8	17
Total Interviews Conducted	4	7	8	19

The usability/cognitive interviews began with the testing of the 2007 GSS Web Survey. All participants were asked to complete this portion of the interview, however one UR was unable to complete the 2007 GSS Web Survey portion of the interview because he did not have access to the Internet in the meeting room. Therefore, a total of 18 participants (4 SCs, 6 URs, and 8 SC/URs) completed the 2007 GSS Web Survey usability/cognitive interview. The two participants who were SCs and who assigned respondents for all units, only completed Part 1 of the GSS Web survey. The six URs completed Part 2 of the GSS Web survey. The eight SC/URs and 2 SCs who entered data for the units, completed both Part 1 and Part 2 of the GSS Web survey.

Next, participants who were unit respondents (SC/URs or URs) were asked to share their input on proposed changes for the 2008 GSS. A total of 17 participants (2 SCs, 8 SC/URs and 7 URs) were interviewed about the proposed 2008 changes. It is important to note, however, that only 6 of these respondents were knowledgeable about the postdoc portion of the survey.

While the majority of the interviews involved two survey methodologists and one respondent, two interviews had multiple respondents. The methodologists met with three people at the Georgia Institute of Technology (Georgia Tech) and two people at University of Washington, School of . At Georgia Tech, the person who filled out the 2007 GSS was new so she was accompanied by the person who had been completing the GSS in previous years, as well as someone who had assisted them both in collecting information for the survey. At the University of Washington, one of the participants had completed Part 1 and had completed the whole survey in past years and the other participant completed Part 2 for the 2007 GSS.

Table 7 identifies the 19 participants by institution, school, title, and GSS coordinator type. As stated before, in some cases there was more than one respondent. For these cases, the title of the recruited coordinator is listed first in the table, followed by the title of the person who assisted with the interview.

Table 7. Usability Test Participants

ID	Institution	School	Title/Office	Coordinator Type
1	University of Minnesota	School of Med	Office of Recruiting for Biomedical Sciences	SC
2	Mayo Graduate School	Grad School	Office of the Education Coordinator	SC-UR
3	Mayo Graduate School	Grad School	Office of Postdoc Coordination	UR
4	University of Minnesota	Grad School	Office of MIS	SC
5	University of Minnesota	Grad School	Office of Applied Plant Sciences	UR
6	Walden University	Grad School	Office of Institutional Researcht	SC-UR

7	Emory University	Grad School	Office of the President	SC-UR
8	Georgia Institute of Technology	Grad School	Office of Institutional Research	SC-UR
9	Georgia State University	Grad School	Chemistry-Lab for Biology and Chemistry	UR
10	Morehouse School of Medicine	Medical Schl	Office of the Graduate Schoolr	SC-UR
11	University of Washington	Public Health	Office of Biostatistics	UR
12	University of Washington	Grad School	Office of the Graduate Dean	SC
13	Oregon State University	Grad School	Office of Institutional Research	SC-UR
14	University of Oregon	Grad School	Office of Graduate Student Affairs	SC
15	Oregon Graduate Institute	Grad School	Office of the Graduate Dean	SC-UR
16	Portland State	Grad School	Office Institutional Research	SC-UR
17	University of Idaho	Grad School	Department of Chemistry	UR
18	Washington State University	Grad School	Department of Computer Science	UR
19	Washington State University	Grad School	Department of Biological Systems Engineering	UR

Results and Conclusions

Overall, participants reported very few problems or difficulties with the 2007 GSS web survey. Most of the changes made to Part 1 of the survey were viewed positively by participants. Participants were favorable to the extended deadline for completing the survey. In addition, none of the participants had a difficult time with the survey being separated into parts. Although, one participant did initially overlook the Part 1 deadline and an additional two participants indicated that the unit listing was not accurate when they submitted it by the deadline. However, both of these participants were SC/URs who provided all of the data themselves. They were also easily able to go back to edit the Part 1 data. Of the participants who had other coordinators at their institutions, all thought that it was helpful to list the other coordinators at their site. Participants found the revised crosswalks, especially the GSS-CIP crosswalk to be extremely valuable in updating their unit listing. None of the participants reported any difficult updating their unit listing, although one participant did initially have some trouble finding the unit listing table at first.

The only section in Part 1 that was somewhat difficult for coordinators was updating and contacting unit respondents. While only two of the test participants assigned all of the units to respondents, both of these respondents indicated that they wanted to be able to download a copy of the contact list to make it easier to send emails offline. One participant, in particular, was frustrated that he could only send IDs and passwords through the system and could not email the

contacts in any other capacity. Another three participants had contacts for all units, but did not enter them into the GSS system because they did not want to email them. Either they wanted the unit respondents to provide the data offline or they just wanted the contacts listed as back ups. To improve the process for future years, coordinators should be able to identify contacts and then indicate whether they will be unit respondents, who complete the survey online, or just alternate contacts. The contact list should also be downloadable to Word or Excel to help coordinators manage their contacts off line.

Most participants did not receive many warnings or errors during Part 1 except for not confirming all units in their listing. None of the participants who received a pop-up error message were correctly able to understand what the problem was and resolve it. We recommend that instead of using a pop-up box, that an error message should appear on the page, but that respondents can navigate away from the page if they choose. If they have not resolved the error by the time they are ready to submit, they will be notified on the error on the Submit Part 1 data page.

The majority of changes made to Part 2 of the survey were also well received. Participants were very favorable of the “View and Submit Data” page, treating it is a road map for the survey. Participants liked the links and buttons that were available on each page, but did not always notice them. We recommend moving the calculate totals button into the table (where the save and refresh button was located in 2006). Because the tables are so large, the calculate totals button was often not visible while participants were entering data. Moving the button to within the table will ensure that it is more noticeable to respondents. Similarly participants were confused by the various options for printing data. They did not understand the difference between “printer-friendly table,” “print worksheet,” and “print data.” In addition the last two options, which were located in the menu bar, frequently went unnoticed by participants. We recommend that the print functions should be explained to respondents at the beginning of the survey and that the “print data” and “print worksheet” options be removed from the menu bar on the top of the screen and placed under the navigation bar. In addition, “download a copy of your survey data” should be placed in a similar location, and it’s purpose explained to participants.

Overall, participants found the warning and error messages to be helpful. The exception was one participant who did not understand what the error numbers in the cells were and another participant who did not notice the explanation of the errors on the survey page. In general, participants felt more favorable towards the error messages than they did the warning messages. When participants received error messages, they commented how helpful they were and how it let them realize what the problems were and that they found them very useful. Participants were more indifferent towards the warning messages noting that they were some that they ignored or just used boilerplate text to respond to. One participant liked receiving the warning messages, but wanted a way to indicate when she had resolved them. Other participants echoed this sentiment indicating that it was good that it flagged it, but that it then became hard to know what was a new warning message and what was a message they had seen, and resolved, already.

Following the review of the 2007 GSS web survey, participants were shown some changes that are being considered for 2008 and future years. Most of these changes such as eliminating NIH Teaching Assistantships, counting students doing their dissertation abroad, and including additional rows for men were well received by participants. While most participants were favorable about including additional rows in the tables for men, three participants were concerned that as a result they may have to edit their programs that pull the race/ethnicity and funding data automatically from a central database. RTI recommends adding the rows because it will improve reporting for most participants, and reduce confusion when errors occur. In addition, all respondents should still be able to complete the revised tables by using subtraction if they cannot get their programs edited in time.

Not all participants were receptive to the clarification in the definition for first-time students. Currently, participants interpreted first-time student in several ways that were inconsistent with the intended definition. For six of these participants, the definition for a first-time student was determined by the university and tracked in a central database. As a result it would be difficult for them to alter their definition to match the GSS. Even though not all institutions will be able to report first-time students according the GSS definition, RTI recommends including the clarified definition of a first-time student (along with some examples of who to include and who to exclude) to ensure that the majority of institutions are reporting using the same standards.

Seven participants out of the 15 who have postdocs at their institutions indicated that they could provide more information about postdocs than what they report now. None of the participants reported being able to provide all of the data that is being considered: race/ethnicity, funding, and U.S. degree. The decision on whether additional questions should be added about postdocs depends on who valuable partial information is. If less than half of the institutions with postdocs can report this information, is it still valuable? Only one participant opposed separating the GSS into a postdoc survey and a graduate student survey. It is possible that separating the postdoc component of the GSS could allow for people who are more knowledgeable about postdocs to report.

All but one participant indicated that it would not be a problem to report separately for Master's students versus PhD students. However, what is meant by a "Master's" student versus a "PhD" student is fuzzy and it varies across institutions and sometimes even within institutions. What would be considered a master's student in one unit may be a PhD in another unit.

D. Testing Alternative Formats

As part of the effort to deconstruct the layout of the current GSS form to make the navigation and instructions more understandable to respondents, NSF will conduct tests of alternative layouts with the GSS respondents. Once the results have been analyzed, NSF will provide OMB with the results and recommendation for the preferred layout for the 2008 GSS.

A summary of the burden hours used for Rounds 1-3 of testing and projected for Round 4 is presented below:

Burden Hours Used Under the GSS-Generic Clearance

Testing Round	Participants	Hours
Round 1 testing (winter 2006/2007)	12	15
Round 2 testing (winter 2007)	24	35
Round 3 testing (Spring 2008)	22	29.5
Round 4 testing (Summer 2008-planned)	12	12

Attachment 6.3
Summary of Meetings with GSS Data Users (November 2007 and January 2008)

Item	Issue/Comment	Response/Suggestion
Masters vs. PhDs	Several users commented that they are "stymied" by the inability to separate masters students from PhD students.	<p>Determine if schools/institutions able to separate masters and PhD students?</p> <p>Ask how many of the students have switched from a masters to PhD program.</p> <p>Ask if students enroll directly into PhD program or get masters first.</p> <p>Count them in PhD years only.</p>
Multi-disciplinary, umbrella, and joint programs	It is unclear how to handle programs that fall into these categories. Worried about undercounting or double counting. How should students pursuing two GSS-relevant degrees simultaneously be counted?	Need to provide instructions or guidelines for handling these types of units.
Students at research centers	Students at research centers may not choose a program or track until after their first year.	Need to provide instructions or guidelines for handling these types of students.
Graduate Student Demographic Data	Want to know if they have ever had a previous graduate experience before.	Unclear how meaningful this would be or if respondents would know this.
	In the tables items 7 and 8 are inconsistent with 1-6.	Consider reformatting tables to be consistent
	The "more than one race" category is confusing.	Suggested changing "Hispanic/Latino" to "and Hispanic/Latino" so that when read with the race heading, "More than one race and Hispanic/Latino." Alternatively the other column should read "but not Hispanic/Latino"
	Increasingly, students are not reporting their race. Concern was expressed about the quality of the race/ethnicity data and the size of the "unknown" category.	Consider imputing race for unknowns.
	What constitutes "new" student. Did not think this was clear. Should transfer graduate students be considered "first-time"? What about students who change the degree they are pursuing but remain at the same institution?	Consider revising definition.
	Why are students enrolled in a U.S. institutions, but doing research in another country excluded? Why exclude students who are studying at a branch or extension of a U.S. institution in a foreign country?	Consider revising instructions.
	One participant indicated that at her academic institution, a PhD was an academic track degree whereas a Doctorate was an industry track degree.	Cognitively test the terms Ph.D. and Doctorate to see if this is an isolated distinction or something that needs to be addressed in the definitions.
	A better explanation of Ph.D. equivalent is needed. One component of a Ph.D. program is completing a dissertation.	Create definition of Ph.D. equivalent and/or provide a complete set of these equivalents rather than examples.

**Attachment 6.3
Summary of Meetings with GSS Data Users (November 2007 and January 2008)**

Item	Issue/Comment	Response/Suggestion
Graduate Student Funding Data	Self-supported students includes Federal loans, which is confusing because self-supported students is under the "Nonfederal sources" columns.	Suggest pulling "self-supported students" out of the Nonfederal sources category and making it its own category. In total, add "including loans."
	Add department of Education as its own category.	Unlikely to add due to space restrictions and limited need for data at this level.
	Other sources of support are not listed. Mentioned were USGS and Homeland Security.	Space restrictions make it difficult to add sources. If these are important, consider if any of the sources currently listed are less important.
	The difference between "institutional support" and "state and local government" is important. Therefore they should not be lumped together. Others stated that institution awards are sometimes state awards so the two could not be separated.	Consider separating the two categories in the tables.
	Under what mechanism of support should tuition waivers go? Other types of support? Tuition waivers are less often the largest source of support than in the past, but there was some pause about tuition waivers "masking" other types of support.	Consider whether tuition waivers should be considered if it is the largest source of support.
	Users would like to know how many students are supported by private foundations.	Consider adding a separate category for "private foundations."
	Definitions for fellowship and traineeship are inaccurate. There are traineeships that don't require work and fellowships that do. A fellowship may be an institutional fellowship that is noncompetitive and does not require work. It was suggested to add institutional fellowships to the definition of fellowships so they are not counted as traineeships.	Revise definitions
	What is the rationalization for separating research assistantships and research traineeships.	Some users find value in this distinction, but it is unclear if respondents can distinguish between the two when reporting.
	Suggested rephrasing the definition of Graduate Research Assistantship: "...responsibilities associated with funding is devoted to research."	Consider revising definition as suggested.
Some desire to have sources of support by race/ethnicity, assuming the quality of the race/ethnicity data is good. It was recognized that this would add burden.	If respondents can provide this information, consider whether to add this distinction given the added burden on respondents.	
Postdocs	Term of postdocs is 1-2 years, not 5-7 years. Sometimes postdocs are called research associates if they have a traineeship.	If able to depart from the joint NIH-NSF definition, consider revising postdoc definition.

Attachment 6.3
Summary of Meetings with GSS Data Users (November 2007 and January 2008)

Item	Issue/Comment	Response/Suggestion
	Would like to collect the same level of detail for postdocs as is collected for graduate students	Unclear if respondents can provide this data for postdocs.
Postdocs (continued)	Would like comparable data on postdocs outside of schools and institutions	Not within the scope of GSS
	Want to have sources of support by citizenship and counts for the number who are US trained vs. foreign trained.	Unclear if respondents can provide this data for postdocs.
	Are the people completing the GSS the best people for providing data about postdocs?	Consider allowing the school coordinator to assign postdoc tables to someone other than the person who completes the graduate student tables.
Nonfaculty researchers	The nonfaculty researcher data was considered extremely important from a workforce perspective, America Competes Act, Carnegie Classification, PhDs who are unable to get a tenure-track position.	Continue collecting the nonfaculty researchers data. Important in its own right.
	Concern about double counting NFRs with joint appointments.	Provide instructions or guidelines for handling these types of researchers.
	Do not define NFRs as "non-tenure track faculty" because GSS does not want instructors and adjunct faculty.	None needed.
	Would like to collect the same level of detail for NFRs as is collected for graduate students	Unclear if respondents can provide this data for NFRs
	Want to know how many postdocs at THAT school/institution end up becoming an NFR at that institution vs. NFRs who came from outside the institution.	Not discussed
Code lists	Consistency of codes is important	
	Some programs are missing from the code list, for example remote sensing and nanotechnology. Others are not in the "right" category (e.g., environmental science, cognitive neuroscience). Some participants were unfamiliar with the way some of the programs were named (e.g., bioinformatics).	
	Need to be more specific about the degrees included. Would like to see more specific fields list.	Consider asking about "degree conferring programs"
Trend data	A concern was expressed with maintaining trend data.	Provide crosswalk and "bridges" in tables.

Attachment 6.4

2008 GSS Pilot Tests

Pilot Test 1: Population coverage and new schools

As part of the GSS redesign effort in 2007, GSS-eligible fields were more fully articulated and the set of GSS-eligible fields of study expanded upon in some areas (i.e., Communication, Family and Consumer Science/Human Science, Multidisciplinary and Architecture). The introduction of a more comprehensive set of GSS-eligible fields presents GSS with the opportunity to ask institutions to assess the completeness of their reporting coverage.

NSF has identified three objectives for this activity. The first is to determine if all of the GSS-eligible fields that were more fully articulated¹ in the 2007 are accounted for by the School Coordinator(s) currently assigned. If it is found that a given institution is not reporting data for some GSS-eligible fields, NSF's second goal is to learn if the current School Coordinator(s) would be able to provide that data themselves or assign unit respondents to do so. In the event that the current School Coordinators are unable to take on the additional burden, NSF's third objective is to explore approaches for identifying a new School Coordinator for these unrepresented GSS-eligible fields.

To meet NSF's goals for this activity, a pilot test will be conducted with 40 institutions. The additional criterion for selection for this pilot test would be an apparent under-coverage of GSS-eligible fields at the institution.

There are two different aspects of under-coverage NSF proposes to investigate. The first is under-coverage of fields which were GSS-eligible in 2007 but were not accounted for in the 2007 data collection. Secondly, some institutions will be selected into the pilot test sample because they have fields of study in business, social work, or education that are being considered eligible programs for purposes of the pilot test. Because these fields are often housed in a dedicated school, they may present special challenges. The current School Coordinators who are often positioned in a graduate school or a medical school may be less able to coordinate a data collection effort in a school that is a distinct entity from their own.

NSF proposes sending advance letters to School Coordinators explaining the goals of the pilot study and requesting a telephone interview. Appointments will be arranged by telephone follow-up. Prior to the interview, NSF will send the School Coordinator a profile of his or her institution's GSS-eligible fields and the units reported in the 2007

¹ Communication and family science are examples. Others examples include fields that were previously not spelled out such as International Agriculture and Cell Physiology. For a complete list of program titles and codes added to the GSS in 2007, see Attachment 1.5: Revisions to the 2007 GSS Codes and Programs.

GSS. If the School Coordinator indicates during the telephone interview that he or she is unable to coordinate the effort for the unreported units, NSF will ask if he or she knows of someone who could serve in that role. NSF would then follow up on this lead. In the event that NSF is unable to successfully identify a School Coordinator using these outlined procedures, NSF may contact an individual in the institution's President's office.

The survey contractor will prepare for NSF's review a complete pilot test plan, criteria for selecting institutions, letters and scripts, and sample institution profiles. Revised versions of these products which incorporate NSF's feedback will be finalized before recruiting for the pilot test begins.

The data for GSS-eligible fields collected as a result of these activities (not including the education, social work, and business programs being tested) will be included in the 2008 GSS data reports. If this pilot test is successful and financial resources are available, NSF will start to survey the fields of business, education and social work in the 2009 GSS survey.

Pilot Test 2: Newly Eligible Institutions

The GSS frame updating research has identified approximately 600 institutions that have not participated in the GSS data collection to date but may be eligible to do so. The ultimate goal is to determine which institutions are eligible for GSS and to include these eligible institutions in the GSS to improve coverage of institutions and thereby increase the accuracy of the data collected and reported. Recruiting these institutions will present some unique and as yet unknown challenges. A pilot study will be conducted in which 80 institutions will be recruited to participate in the 2008 GSS data collection cycle. Through this pilot test, we will learn the best way to identify the appropriate contact at the institution, gain the support of that individual, assist that individual in identifying the scope of the GSS at his or her institution, gain the cooperation of individuals nominated to serve as School Coordinators, help these new School Coordinators identify GSS-eligible units, and assist them during data collection. The data collected from these newly eligible institutions will be included in the 2008 GSS data reports. The lessons learned from this pilot test will inform procedures and system development for future rounds of the GSS. If the pilot test is successful and financial resources are available, NSF will add more GSS-eligible institutions in the 2009 GSS survey.

Attachment 6.5

Process for Determining GSS-relevant CIP Codes

The GSS subset of Classification of Instructional Programs (CIP) codes included are those in science, engineering, and health disciplines that appear to have a research-oriented basis and are not primarily practitioner-preparation or “how-to” programs. Excluded are dental (60.01), medical (60.02), and veterinary (60.03) residency programs and first-professional degree programs (e.g., PharmD, MD, DO, DVM, DDS).

The CIP analysis was based on multiple sources, but primarily the 2006 IPEDS Completions data which were collected in the fall of 2006, in conjunction with an analysis of the Student and Exchange Visitor Information System (SEVIS) database done by Survey Sciences Group for SRS’s Postdoc Data Project; the academic institutions’ Websites describing specific programs; and general Web searches to help make a determination for each CIP. When pulling the data from IPEDS, the frequency of each CIP by award level (associate’s, bachelor’s, master’s, doctorates, and first-professional) was examined. These frequencies from IPEDS provided an idea of how a specific CIP is awarded at different degree levels. From the IPEDS data, a sample of institutions associated with each specific CIP was drawn. The academic institutions’ Websites were then examined for a description of the graduate program to help determine the CIP’s relevance to GSS. When looking at the Websites, key terms and descriptions of the programs such as: thesis, research oriented, or leads to positions in research areas were looked for. The determination for inclusion or exclusion of a particular CIP as GSS-eligible was based on the cumulative information obtained.

The rules used to determine the relevance to GSS of “borderline” CIPs were similar but more rigorous and involved a larger sample of graduate programs. In cases where a strong case could not be made for recommending a CIP to be included or excluded, it was set aside for further consultation with NSF/NIH.

When a CIP is both a field and an occupation, the decision was to exclude the practitioner-only degree programs but include master’s degree and doctoral degree programs that have theses or dissertations that are research-oriented. There were a few cases within the same CIP where the master’s level appears to be a practitioner-preparation program, and the PhD level is research-oriented. Dental Hygiene/Hygienist is a good example of where this occurs. Therefore, a parenthetical note in the CIP title was added to exclude Master’s-level awards for these specific CIPs.

So that information regarding postdoctoral appointees and nonfaculty researchers in clinical areas may be included, the GSS-relevant instructional program code frame has been supplemented to include clinical disciplines that otherwise are not a part of the GSS-relevant CIP list. These supplemental codes include NIH fields of training and additional areas as approved by NIH and NSF.