STUDY SERIES
(Survey Methodology #2011-01)

A Usability and Eye-Tracking Evaluation of Four Versions of the Online National Survey of College Graduates (NSCG): Iteration 2

Jennifer C. Romano
Jennifer M. Chen

Center for Survey Measurement
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

A Usability and Eye-Tracking Evaluation of Four Versions of the Online National Survey of College Graduates (NSCG): Iteration 2

Abstract:
The National Survey for College Graduates (NSCG) is an online Web survey that collects education and job information from college graduates with at least a bachelor's in a science or engineering field. In February and March 2010, the Statistical Research Division (SRD) evaluated the usability of the revised NSCG online survey. This study followed an initial round of testing that was conducted in October and November 2009 and subsequent changes that were made to the survey following usability recommendations. This testing evaluated the success and satisfaction of 30 participants who attempted to complete the Web-based survey, developed by the Special Surveys Branch (SSB) of the Demographic Surveys Division (DSD), in collaboration with the National Science Foundation (NSF). Usability testing revealed that it was more efficient to display the long list of occupation options in two columns rather than one column, and to display the 'Next' button to the right of the 'Previous' button rather than vice versa. Other usability issues are reported, as well as the methods and findings of this usability evaluation.

Key Words:
Web survey, Internet form, Eye tracking, Usability

Executive Summary

In February and March 2010, the Statistical Research Division (SRD) Usability staff evaluated the usability of the National Survey of College Graduates (NSCG) Web-based survey. The testing evaluated the success and satisfaction of 30 participants with the survey developed by the Special Surveys Branch (SSB) of the Demographic Surveys Division (DSD), in collaboration with the National Science Foundation (NSF). Testing took place at the Census Bureau's Usability Laboratory in Suitland, MD.

Purpose. The primary purpose of this usability test was to identify elements of the user-interface design that were problematic and led to ineffective, inefficient, and unsatisfying experiences for people completing the survey. We tested four versions of the instrument to assess the usability of the 'Next' and 'Previous' buttons based on their placement on the page (right or left side) and the job code collection format (one-column versus two-column).

Method. Thirty people, with a mean age of 46 years (range: 25-75 years), participated in the study. Four versions of the survey were evaluated. Participants read a log-in letter and followed the instructions to access the online survey. While completing the survey, they were asked to verbalize what they were thinking about. For example, participants were encouraged to voice questions or opinions about the survey. If at any time the participant became quiet, the test administrator reminded the participant to think aloud. After the participant completed the survey, he or she completed a Satisfaction Questionnaire and answered debriefing questions. Eye-tracking data were collected. Overall, each session lasted about one hour.

Results. Testing identified three high-priority, four medium-priority, and three low-priority usability issues. The following are highlights of the three high-priority issues:

1. Participants did not read the text on the log-in screen. Eye tracking demonstrated that most users did not read the text on the log-in screen but read the text on the Introduction screen (which follows log in). During testing and debriefing, participants said that they wanted to know why the survey was being conducted and why some questions were being asked. It is therefore important to include any pertinent information on the Introduction screen instead of, or in addition to, the log-in screen.

2. It was counter-intuitive for the 'Previous' button to be on the right and the 'Next' button to be on the left side of the screen. During debriefing, when participants were asked what they thought of the 'Next' and 'Previous' buttons, many said that the buttons were fine, regardless of their location. However, it appears that when the buttons were ordered with 'Next' to the left of 'Previous,' a few participants expressed strong dissatisfaction with the button order. No participants complained about the button order when the 'Next' button was to the right of the 'Previous' button.

3. Having a long list of job or education options was not preferred by participants. Of the 25 participants who said that they had a preference, 19 said they preferred a two-column format, and six said they preferred a one-column format. Eye-tracking data showed no difference in the efficiency with which people looked at the second half of the options.

Table of Contents

List of Figures

List of Tables

A Usability and Eye-Tracking Evaluation of Four Versions of the Online National Survey of College Graduates (NSCG): Iteration 2

1.0 Introduction & Background

The user interface is an important element to the design of a Web-based survey. For the instrument to be successful, the interface must meet the needs of its users in an efficient, effective, and satisfying way. It is the job of the interface to provide cues and affordances that allow users to get started quickly and to find what they are looking for with ease, and the design of the instrument must support users' navigation. For a survey, success is defined in terms of being able to successfully respond to questions in the survey. In addition, the overall experience of completing the survey should be satisfying for users.

This report specifies the methods and materials that the Statistical Research Division (SRD) used to evaluate the usability of the National Survey of College Graduates (NSCG) Web-based survey. The NSCG is a data collection instrument that collects education and job information from respondents who have received bachelor's degrees in science or engineering fields from American schools or abroad. It is currently available as a paper-and-pencil instrument (PAPI), and the non-response follow-up is available as a computer-assisted telephone interview (CATI). The web-based instrument tested here is expected to be available in October 2010.

This usability test evaluated the success and satisfaction of test participants with the online survey developed by the Special Surveys Branch (SSB) of the Demographic Surveys Division (DSD), in collaboration with the National Science Foundation (NSF). This usability test followed an initial round of testing that occurred October 29 to November 19, 2009 (Romano and Chen, 2010) on an initial version of the instrument. In Iteration 1 testing, eight people participated. Usability issues were discovered, and the SSB team made changes to the instrument.

In this round of testing, we tested four versions of the NSCG online survey. The different versions allowed us to assess the effects of the location of the 'Previous' and 'Next' buttons and a long list of response options displayed in one long list versus two lists side-by-side on user performance and satisfaction. Participants completed the online survey during testing (see Figure 1 for selected screen shots of the instrument). Findings are provided in this report to inform the site sponsor and designer(s) on areas of satisfaction, as well as areas where the participants struggled while completing the NSCG online survey.

Thirty people participated in this usability study between February 17 and March 19, 2010. Members of SRD's Usability Lab prepared and disseminated a brief report of major findings and recommendations to DSD on March 30, 2010. The DSD team responded to the findings and recommendations on March 30, 2009. This report documents the results of testing and the responses from the DSD team. This report also compares the findings from Iteration 1 (Romano and Chen, 2010) to the present findings.
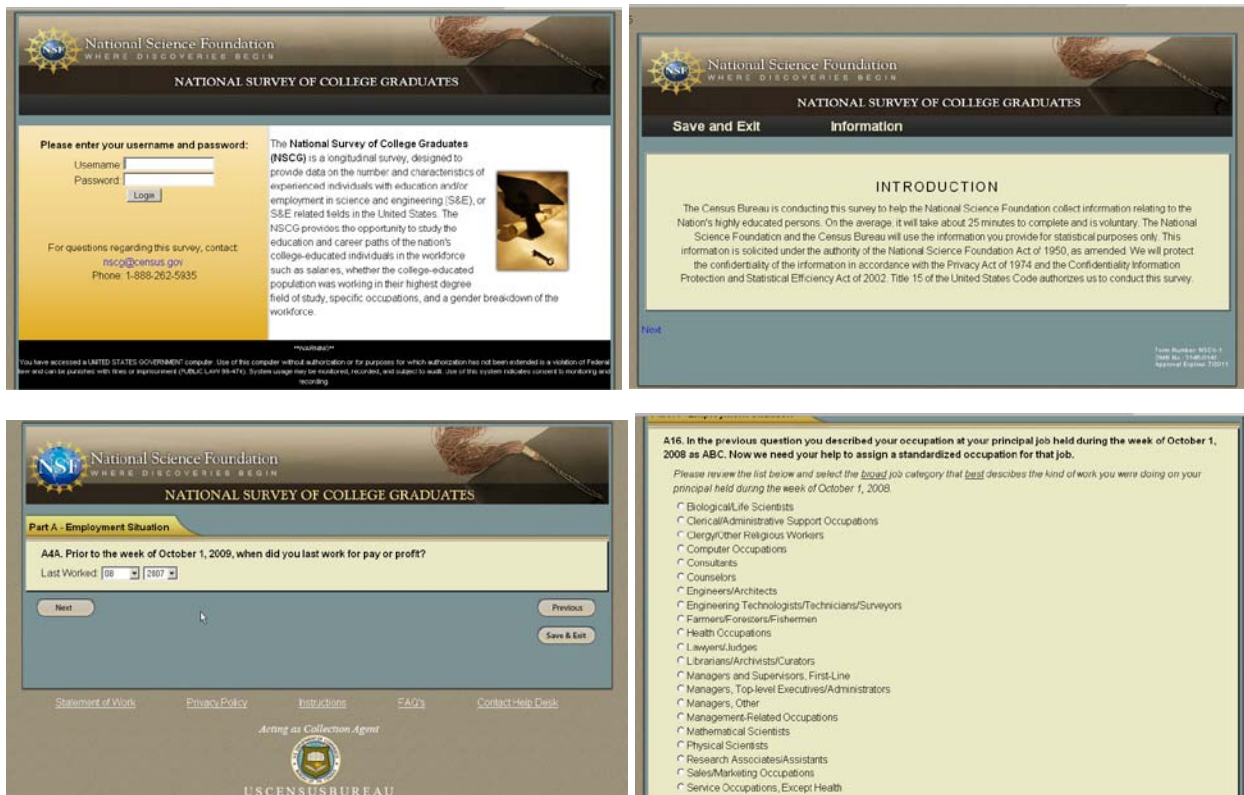
Figure 1. Selected screen shots of the NSCG online survey.

1.1 Purpose

The primary purpose of this evaluation was to identify elements of the user-interface design that were problematic and led to ineffective, inefficient, and unsatisfying experiences for people completing the survey. Several features of the survey were of particular interest. They were:

1. The occupation code and education code data collection method
2. The order of the 'Previous' and 'Next' buttons

1.2 Usability Goals

We defined the usability goals for this study in two categories: user accuracy and satisfaction. In this study, these goals reflect the extent to which the user interface was expected to support user performance and satisfaction. Typically in usability testing, an efficiency goal is established; however, in this usability study, the efficiency measure would be uninformative.

Goal 1, Accuracy: To achieve a high level of accuracy in recording the answers on the NSCG online survey. The respondent should successfully respond to 100% of the answers on the survey by using the response options correctly and without encountering unsolvable problems. Note that this study did not assess whether the participants chose the *correct* response options, but instead it assessed whether participants could use the form. The user should be able to correctly navigate through the entire questionnaire.

Goal 2, Satisfaction: To experience a moderate to high level of satisfaction from experience using the NSCG online survey. The overall mean of the Satisfaction Questionnaire ratings should be well above the mid-point (5 on a nine-point scale, where 1 is the lowest rating and 9 is the highest rating). The same should be true for the individual Satisfaction Questionnaire items.

2.0 Method

2.1 Versions of the Survey

In order to assess the effects of the locations of the Previous and Next buttons and to assess the effects of occupation code data-collection method (one column vs. two columns), the Special Surveys Branch team created four versions of the NSCG survey. They were:

NP1 'Next' button on left, 1-column job code	**NP2** 'Next' button on left, 2-column job code
PN1 'Previous' button on left, 1-column job code	**PN2** 'Previous' button on left, 2-column job code

Only one version was tested at a time, and each version was tested for approximately one week. Each version was tested on eight people, except for Version PN1, which was tested on six people, because two scheduled participants cancelled their sessions.

2.2 Participants and Observers

Thirty people (fourteen males, sixteen females), with a mean age of 46 years (range: 25-75 years), participated in the usability study. Twenty-seven people were recruited externally through a database maintained by the Usability Lab, through advertisements placed in a local newspaper and through a local list serve. Three people were recruited internally by the Usability Team. See Table 1 for overall demographics and Table 7 in Appendix A for demographics by group (i.e., survey version). All participants reported having at least one year of prior Internet and computer experience and prior knowledge of how to navigate a Web site (see Table 8 in Appendix B). Participants qualified for taking the NSCG by having earned a bachelor's degree and currently residing in the United States.

Table 1. Participant Demographics

Gender	N	Age	N	Education	N	Race	N
Male	14	≤ 30	8	Bachelor's	21	African American**	15
Female	16	31-45	7	Master's	6	Asian**	2
		46-60	10	Ph.D.	3	American Indian or Alaska Native***	1
		> 60	5			White***	14
Mean		**45.87 years***					

*The mean age was calculated from the exact values for each participant. The exact self-reported values were placed in ranges in the table to help the reader get an overview of the data.
**One participant marked African American and Asian.
***One participant marked White and American Indian or Alaska Native.

Observers from DSD and Special Sworn Status employee staff from the National Science Foundation (NSF) watched the usability tests on a television screen and computer monitor in a room separate from the participant and test administrator (TA). At the end of each test session, the TA and observers discussed the findings from that session and compared them to findings from other sessions.

2.3 Facilities and Equipment

Testing took place in the Usability Lab (Room 5K512) at the U.S. Census Bureau in Suitland, MD.

2.3.1 Testing Facilities

The participant sat in a 10' x 12' room, facing one-way glass and a wall camera, in front of a standard monitor that was on a table at standard desktop height. During the usability test, the TA sat in the control room on the other side of the one-way glass. The TA and the participant communicated via microphones and speakers.

2.3.2 Computing Environment

The participant's workstation consisted of a Dell personal computer with a Windows XP operating system, a standard keyboard, and a standard mouse with a wheel. The screen resolution was set to 1024 x 768 pixels, and participants used the Internet Explorer browser.

2.3.3 Audio and Video Recording

Video of the application on the test participant's monitor was fed through a PC Video Hyperconverter Gold Scan Converter, mixed in a picture-in-picture format with the camera video, and recorded via a Sony DSR-20 Digital Videocassette Recorder on 124-minute, Sony PDV metal-evaporated digital videocassette tape. One desk and one ceiling microphone near the participant captured the audio recording for the videotape. The audio sources were mixed in a Shure audio system, eliminating feedback, and were then fed to the videocassette recorder.

2.4 Materials

2.4.1 Researcher Script for Usability Session

The TA read some background material and explained several key points about the session. See Appendix C.

2.4.2 Consent Form

Prior to beginning the usability test, the participant completed a consent form. See Appendix D.

2.4.3 NSCG Log-in Letter

The log-in letter contained directions about how to access the online survey and log in. See Appendix E.

2.4.4 Questionnaire on Computer and Internet Experience and Demographics

Prior to the usability test, the participant completed a questionnaire on his/her computer and Internet experience and demographics. See Appendix B.

2.4.5 Satisfaction Questionnaire

Members of the Usability Lab created the Satisfaction Questionnaire, which is loosely based on the Questionnaire for User Interaction Satisfaction (QUIS, Chin, Diehl, & Norman, 1988). In typical usability tests at the Census Bureau, we use 10 to 12 satisfaction items that are tailored to the particular user interface we are evaluating. In this study, the Satisfaction Questionnaire includes 10 items worded for the NSCG online survey. See Appendix F for the complete questionnaire.

2.4.6 Debriefing Questions

After completing all tasks, the participant answered debriefing questions about his/her experience using the NSCG online survey. See Appendix G.

2.5 Procedure

Following security procedures, external participants individually reported to the visitor's entrance at the U.S. Census Bureau Headquarters and were escorted to the Usability Lab. Internal participants met the TA at the Usability Lab. Upon arriving, each participant was seated in the testing room. The TA greeted the participant and read the general introduction. Next, the participant read and signed the consent form. After signing the consent form, the TA showed the log-in letter to the participant and explained that the participant should act as if they had received the letter in the mail and should follow the instructions in the letter to access the survey. The TA then placed the NSCG log-in letter, the Questionnaire on Computer and Internet Experience and Demographics, and Satisfaction Questionnaire on the desk beside the participant and left the testing room. While the TA went to the control room to perform a sound check, the participant completed the Questionnaire on Computer and Internet Experience and Demographics[1]. The TA then began the video recording. The internet browser was pre-set to the Census Bureau home page (www.census.gov), and the TA instructed the participant to read the log-in letter aloud and proceed as they normally would at home.

While completing the survey, the TA encouraged the participants to think aloud and to share their thoughts about the survey. For example, participants were instructed to voice questions and opinions about what they liked and did not like about the survey. The participant's narrative allowed the researchers to gain a greater understanding of how the participant completed the survey and to identify issues with the survey. If at any time the participant became quiet, the TA reminded the participant to think aloud, using prompts such as "What are you thinking?" and "Tell me your thoughts." During the sessions, the TA noted any behaviors that indicated confusion, such as hesitation, backtracking, and comments. After survey completion, the TA asked the participant to complete the Satisfaction Questionnaire.

---

[1] Four participants completed the questionnaire when they were approximately halfway through the survey. Identical to Iteration 1 (Romano & Chen, 2010), at a pre-determined point during the session (following Section A of the survey), the TA interrupted the participant and asked them to exit out of the survey and to fill out the Questionnaire on Computer and Internet Experience and Demographics. They were then asked to return to the Web site to resume completing the partially-completed survey. In Iteration 1, we tested the re-log-in procedure. In Iteration 2, following the first four participants, the sponsor decided to forgo this procedure for the remaining participants because they felt sufficient data on this procedure was collected.

While the participant completed the Satisfaction Questionnaire, the TA met with the observers to see if they had any additional questions for the participant. The TA then returned to the testing room to ask debriefing questions. See Appendix G for the debriefing questionnaire. This debriefing provided an opportunity for a conversational exchange with participants. The TA remained neutral during this time to ensure that they did not influence the participants' reactions to the survey. At the conclusion of the debriefing, the TA stopped the video recording. Overall, each usability session lasted approximately 60 minutes. External participants were paid $40 each.

2.6 Performance Measurement Methods

2.6.1 Accuracy
After each participant completed the survey, the TA rated the completion as a success or a failure. In usability testing, successful completion of a task means that the design supported the user in reaching a goal. Failure means that the design did not support task completion. In this usability study, a success involved the participant answering 100% of the items that they encountered on the survey by using the response options appropriately and without encountering unsolvable problems. If the participant struggled to complete the survey but eventually was able to, this was a success.

2.6.2 Satisfaction
After completing the usability session, each participant completed the tailored ten-item Satisfaction Questionnaire[2]. Participants were asked to rate their overall reaction to the survey by selecting a number from 1 to 9, with 1 being the lowest possible rating and 9 the highest possible rating. Other items on the questionnaire included assessing the screen layouts, the use of terminology on the Web site, the instructions and questions displayed on the screens, the arrangement of information on the screens, and the overall experience of completing the survey. See Appendix F for the complete questionnaire. The Usability Team calculated ranges and means for the various rated attributes of the survey. Prior to usability testing, the DSD team and the Usability Team set satisfaction goals for the overall mean and the individual item means to be well above the mid-point of the nine-point scale (e.g., 5 or above).

2.7 Identifying and Prioritizing Usability Problems
To identify design elements that caused participants to have problems completing the survey, the TA recorded detailed notes during the usability session. When notes were not conclusive, the TA used the videotape recordings from each session to confirm or disconfirm findings. By noting participant behavior and comments, the Usability Team inferred the likely design element(s) that caused participants to experience difficulties. The team then grouped the usability issues into categories based on severity and assigned each problem a priority code, based on its effect on performance. The codes are as follows:

---

[2] As part of another ongoing study assessing the effects of administration mode on satisfaction ratings (Jans, Romano, Ashenfelter, & Krosnick, in prep), nine participants completed the Satisfaction Questionnaire on paper in the same room as the usability test; eight people completed it on paper in an adjacent room; ten people completed it on the same computer as the usability test (in the same room); and three people completed it on a computer in an adjacent room. The sample size of each group is not equal because participants were randomly assigned to each group.

- *High Priority* – These problems have the potential to bring the participant to a standstill. (e.g., Participant was not able to successfully complete the survey because they could not log in, re-log-in after interruption, or select the correct job code.)
- *Medium Priority* – These problems caused some difficulty or confusion, but the participant was able to successfully complete the survey.
- *Low Priority* – These problems caused minor annoyances but did not interfere with the completion of the survey.

3.0 Results

In this section, we discuss the findings from the usability study. We present the qualitative and quantitative data, usability issues, and possible future directions based on the DSD team's responses to the findings. Two-tailed independent *t*-tests were performed on the measures to compare the different versions, except where indicated otherwise.

3.1 Participant Accuracy

All participants were able to successfully complete the NSCG online survey. However, some participants skipped "sensitive" items, such as reporting contact information of friends or relatives and reporting income. One person skipped many of the survey questions. It did not appear that she felt the survey questions were overly hard or sensitive but just did not feel like completing those items.

One reason people may have skipped items is because they did not understand why the data was being collected. During testing and debriefing, participants said that they wanted to know why this survey was being conducted and why it was asking for personal information. Similarly, some participants mentioned that they wondered why certain questions were being asked. This is discussed further in Section 3.6.1.1 of this report.

3.2 Participant Satisfaction

The average satisfaction score across all participants was 7.65 out of 9 (1 = low and 9 = high). See Table 2. This average suggests that users were satisfied with the Web-based survey, overall. The highest rated items were "forward navigation: impossible – easy" and "screen layout: confusing – clear," with average scores across participants of 8.37 and 8.13, respectively. None of the satisfaction score averages were below 7.00, suggesting high satisfaction among users.

Table 2. Mean Satisfaction Questionnaire Scores Across Participants, Across Versions

Version	Satisfaction Questionnaire Item										Mean
	1	2	3	4	5	6	7	8	9	10	
NP1	6.25	7.88	7.13	7.75	8.00	7.25	6.50	7.25	8.25	7.75	**7.36**
NP2	7.00	8.00	7.38	7.13	7.50	6.75	6.88	6.88	7.88	7.50	**7.29**
PN1	8.17	8.67	8.17	8.33	7.83	8.25	7.17	8.00	8.67	8.33	**8.12**
PN2	7.50	8.13	7.63	8.13	7.75	7.86	7.50	7.57	8.75	8.13	**7.94**
Overall mean	**7.17**	**8.13**	**7.53**	**7.80**	**7.77**	**7.41**	**7.00**	**7.38**	**8.37**	**7.89**	**7.65**

When comparing the different versions of the survey, there was no difference in mean satisfaction across the four versions of the survey, $F(3) = 0.82$, $p = 0.50$. The average satisfaction score, in decreasing order, across users was 8.12 for the PN1 version; 7.94 for the

PN2 version; 7.36 for the NP1 version; and 7.29 for the NP2 version. See Table 2. However, when the PN versions were grouped together and the NP versions were grouped together, there was a trend for users to be more satisfied with the PN versions, $F(1) = 2.55$, $p = 0.12$. See Table 3. When the 1-column versions were grouped together and the 2-column versions were grouped together, there was no difference in mean satisfaction across users, $F(1) = 0.02$, $p = 0.89$. See Appendix F for satisfaction scores by version and for the complete satisfaction questionnaire.

Table 3. Mean Satisfaction Collapsed Across Versions

Condition Type	Satisfaction
NP	7.34
PN	8.01
1 column	7.72
2 column	7.59

The mean across participants for Questions 1, 3, 6, 7, 9, and 10 were lower for Versions NP1 and NP2 compared to Versions PN1 and PN2, as shown in Table 2. Those questions were:
- #1: "Overall reaction to the site: terrible – wonderful" $t(28) = 2.10$, $p < 0.05$[3]
- #4: "Information displayed on the screens: inadequate – adequate" $t(28) = 1.92$, $p = 0.07$
- #6: "Arrangement of information on the screens: illogical – logical" $t(24) = 0.70$, $p = 0.19$
- #7: "Questions can be answered in a straight-forward manner: never – always" $t(28) = 1.14$, $p = 0.27$
- #9: "Forward navigation: impossible – easy" $t(28) = 1.58$, $p = 0.13$
- #10: "Overall experience of completing the survey: difficult – easy" $t(28) = 1.16$, $p = 0.26$.

Consistent with Iteration 1 findings (Romano & Chen, 2010), the low satisfaction scores of items #6 and #9 may reflect participants' dissatisfaction with the location of the 'Next' button when it was to the left of 'Previous.' This is discussed further in Section 3.6.1.2 below.

3.4 Positive Findings
- All participants completed the survey without being brought to a standstill.
- All participants said that they thought the design and layout of the site was attractive.
- Participants said that they felt the survey questions were straightforward and that the survey had a logical progression.
- Several participants remarked that the survey was not too long.
- Participants read the text on the Introduction page.
- Participants said that, for the most part, selecting their job code was not difficult.
- All participants, except one, selected the correct job category.[4]

---

[3] The reported statistics are for Versions NP versus Versions PN (collapsed).
[4] The participant who did not select the correct category was self-employed and did "odd jobs." For the job selection, he did not make a selection for self-employed. Rather, he described a specific job he did and reported the income from that job. He proceeded to do the same for a second odd job, and then inquired if he should continue to report many such jobs. The test administrator told him that what he had provided was sufficient and instructed him to move on with the survey.

- The revised (from Iteration 1) 'Other' text box worked well for users. They used the box to type in appropriate information, and none of the participants reached the character limit.
- Participants said that the large font size and navigation buttons were easy to use.

3.5 Interpreting Eye-Tracking Results

3.5.1. Heat Maps

In eye tracking, a fixation is an instant where the eyes are relatively still (Poole & Ball, 2005). Heat maps take fixation data and convert them into a visual format where the fixations are represented by colors. Heat maps range in color from green (indicating few fixation points) to red (indicating many fixation points) and can be generated for individual participants or for a group of people to show the average number of fixations on a screen.

3.5.2. Gaze Plots

A gaze plot represents the total number of fixations in a uniquely defined area. Gaze plots indicate which areas are getting the most attention (Poole & Ball, 2005) by showing the order and duration of fixations. Circles represent the fixations, and the size of the circle represents the duration. The numbers in the circles show the sequence of fixations, such that "1" represents the first recorded fixation, "2" represents the second recorded fixation, and so on. Gaze plots can only be generated for one participant at a time.

3.6 Usability Issues and Explanations

Explanations for performance deficits and descriptions of usability violations are discussed in this section. High-priority usability issues have the potential to bring users to a stand-still. Medium and low-priority issues may lead to inefficiency and/or user frustration but generally do not interfere with task completion. Testing identified three high-priority usability issues, four medium-priority usability issues, and three low-priority usability issues, which are detailed below. Addressing these issues should result in improvement in the users' performance and satisfaction with the NSCG Web-based survey.

3.6.1 High Priority Usability Issues

3.6.1.1 Participants did not read the text on the log-in screen.

On the log-in screen, most of the participants did not read the text to the right of the log-in entry boxes. See the upper left panel of Figure 1 for a screen shot of the log-in screen, and the left panel of Figure 2 for a screen shot with an eye-tracking heat map overlay. Here we collapsed hot spots across participants to form a heat map for a mean image that displays the average of all hot spots, for all the participants.

The heat map displayed in the left panel of Figure 2 and the gaze plots in Appendix H demonstrate that most people do not look at the text on the log-in screen. They do, however, read the text on the Introduction screen, which is the first page to appear following successful log in, as shown in the heat map in the right panel of Figure 2 and the gaze plots in Appendix I.

Figure 2. **Heat map of the log-in screen (left panel) and the Introduction screen (right panel), across all participants, demonstrating that participants did not look at the text on the log-in page, but they looked at the text on the Introduction page.**

During testing and debriefing, participants said that they wanted to know why this survey was being conducted and why it was asking for personal information (both of which are explained on the log-in screen). This suggests that any important information should be displayed on the Introduction screen, either in place of the log-in screen, or in addition to it. Providing this information in a usable way (i.e., in a way that users will read it) will likely lead to higher response rate and lower drop-out rate, as respondents will be informed about the importance and uses of the data.

Similarly, some participants mentioned that they wondered why certain questions were being asked. Particularly, they felt it strange that the survey used such specific dates and time frames. Adding in some context to why questions are asked may make participants feel more comfortable answering the questions. Some questions already have a brief explanation accompanying them (e.g., the disability question). These explanations could also be given on the Introduction screen. Alternatively, a link with the heading "Why is this being asked?" or an information icon that when clicked provides a brief rationale, could also be implemented.

Sponsor Response: This may be an artifact of the usability session. Participants were asked to read the letter and then to log in, which may have taken their attention as the TA told them already what the survey was about. However, as there is no way to test this, careful consideration will be paid to what information should be displayed on the Web survey log-in screen and Introduction screen.

3.6.1.2 It was counter-intuitive for the 'Previous' button to be on the right and the 'Next' button to be on the left side of the screen.

Figure 3 and Figure 4 display eye-tracking data for the different versions of the survey. Figure 3 displays gaze plots from Versions NP1 and NP2, in which 'Next' is on the left and 'Previous' is on the right, and Figure 4 displays gaze plots from Versions PN1 and PN2, in which 'Previous'

is on the left and 'Next' is on the right. The screen depicted is one that all participants encountered at the beginning of the session. The item on this screen asked participants to verify the college they received their Bachelor's degree from.

During debriefing, when the TA asked participants what they thought of the 'Next' and 'Previous' buttons, many said that the buttons were fine, regardless of their location. However, as shown in Figure 3, when using the NP1 and NP2 versions, many participants looked at the 'Previous' button on the right side of the page, suggesting that it may be distracting to them.
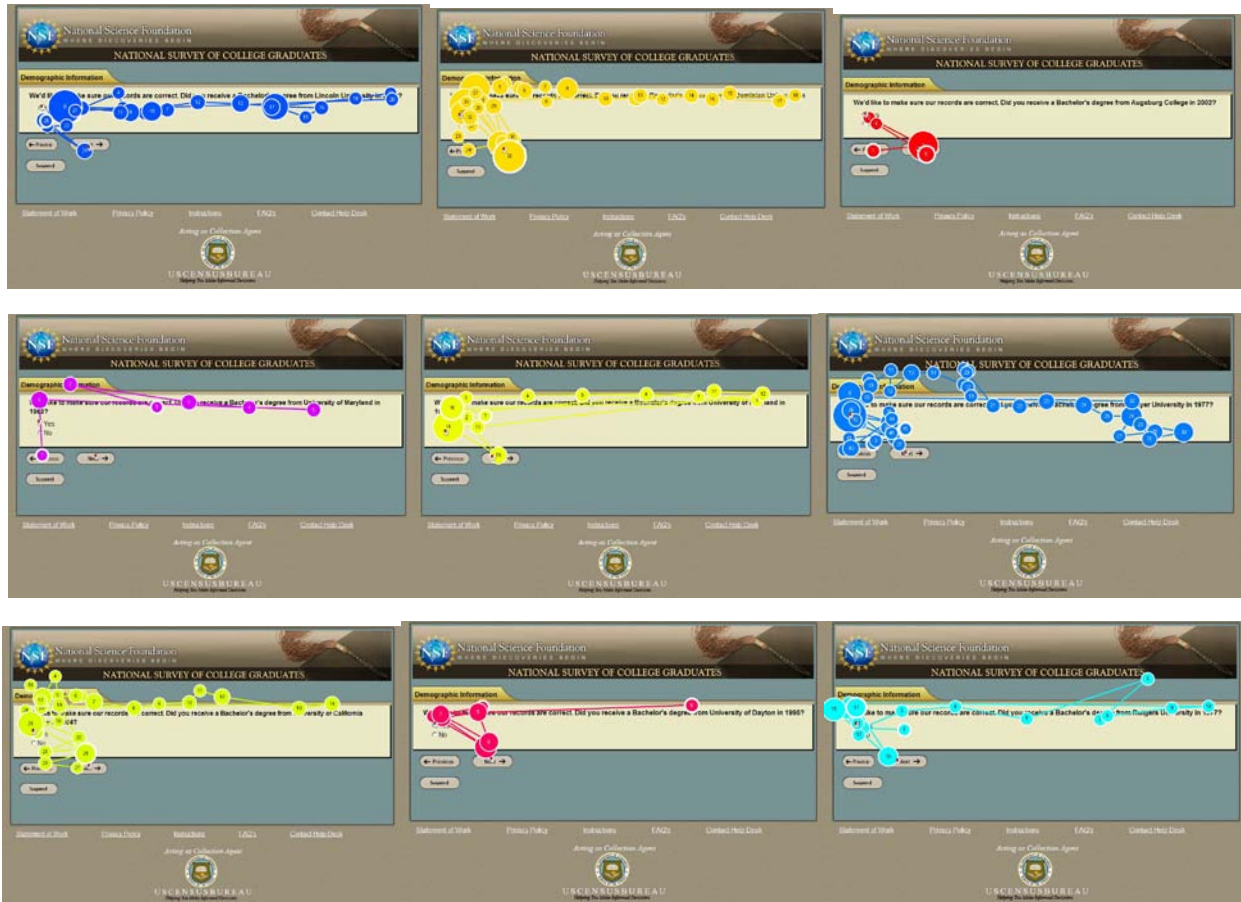

Figure 3. **Screen shots displaying gaze plots for participants who responded to the versions in which the Next button was on the left and the Previous button was on the right.**

Many participants said that it was counter-intuitive to have the 'Previous' button on the right. One participant said that she disliked the "buttons being flipped" although she liked the look and

size of the buttons. Another participant said that having the Next button on the left side "really irritated" him. Another said that the order of the buttons was "opposite of what most people would design."

As shown in Figure 4, when using the PN1 and PN2 versions, participants often looked at both the Previous and Next buttons[5]. However, no one explicitly claimed that the location of the buttons was problematic. One participant said that the buttons looked "pretty standard, like what you would typically see on Web sites." Another said the location was "logical."

Figure 4. **Screen shots displaying gaze plots for participants who responded to the versions in which the Previous button was on the left and the Next button was on the right.**

Overall, it appears that the PN condition elicited a few participants to express strong dissatisfaction with the button order while the NP condition did not elicit any strong opinions.

Sponsor Response: The 'Previous' button will likely be to the left of the 'Next' button in the final version of the Web-based survey.

---

[5] Note that the location of the right button in the two versions is slightly different.

3.6.1.3 Having a long list of occupation items in one scrolling list is not efficient.

Table 4 displays the average amount of time that elapsed before participants first looked at the two halves (first half, second half) of the job code occupation response list as well as the number of fixations on each half of list. The table presents data from one-column versions and two-columns versions of the Web survey separately. Figure 5 displays screen shots of the two versions. Appendix J displays individual participant gaze plots for the one-column versions, and Appendix K displays gaze plots for the two-column versions.

Table 4. Eye-Tracking Data for the Occupation Item: Mean Across Participants

Version	Total seconds elapsed before first fixation		Total number of fixations	
	First half of list	Second half of list	First half of list	Second half of list
2 column	9.92	18.30	35.38	28.85
1 column	6.60	20.64	36.75	17.83

As shown in Table 4, and as is typical, participants looked at the first half of the list before the second half of the list in both the one-column version, $t(21) = 3.35$, $p < 0.005$, and two-column version, $t(23) = 1.50$, $p = 0.15$. There was no difference in the total number of fixations for the first half of the list versus the second half of the list for both the one-column version, $t(22) = 1.29$, $p = 0.21$, and the two-column version, $t(24) = 0.56$, $p = 0.58$.
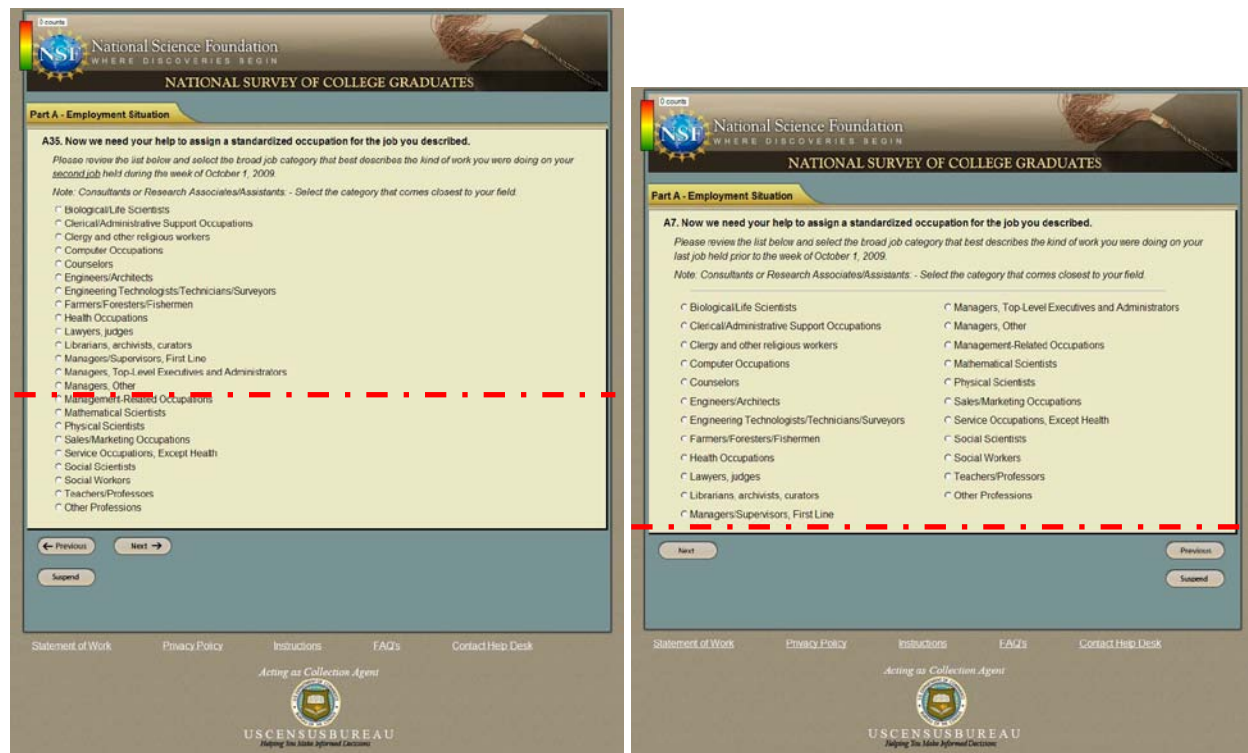
Figure 5. Screen shots of the occupation items in one column (left panel) and in two columns (right panel). Red dotted lines denote the "fold" of the page, such that users had to scroll to see what was below the fold.

While it appears that people looked at the second half of the list more quickly and more often when it was in two-column format compared to the one-column format, these differences were not statistically significant, $t(21) = 0.04$, $p = 0.69$, $t(23) = 1.19$, $p = 0.24$, respectively.

The total time it took participants to complete the first occupation item was calculated. On average, participants spent 44 seconds answering the item when the occupation types were listed in two columns and 48 seconds when they were listed in one column. For a complete list of times for each participant, see Table 5. Thus, it appears that people were more efficient and answered the item faster when the list was displayed in two columns; however, this difference was not statistically significant, $t(28) = -0.49$, $p = 0.63$.

Table 5. Total Time (in Seconds) to Complete the Occupation Item

1 column		2 column	
Participant	**Time**	**Participant**	**Time**
16	97	1	14
17	19	2	103
18	39	3	32
19	43	4	20
20	29	5	23
21	32	6	82
22	40	7	27
23	82	8	21
24	77	9	31
25	54	10	50
26	18	11	45
27	36	12	18
30	82	13	54
31	36	14	78
		15	80
		29	27
Mean	**49 seconds**		**44 seconds**

During debriefing, the TA asked participants if they preferred the occupation list to be displayed in one column or two. Of the 25 participants who said that they had a preference, 19 said they preferred two columns, and six said they preferred one column. All of the six who said they preferred one column had received the one-column version of the survey, so it is impossible to definitively determine whether they would have preferred a two-column version if they had seen it. However, of the 25 participants who said they preferred two columns, seven had received the one-column version. Even though they had not been exposed to the two-column version, they still said they would have preferred the list to be in two columns instead of one. Some of the reasons they gave for wanting the list in two columns include: having a strong preference not to scroll, wanting to see everything at once, having the ability to compare all options at once, being easier to "look through" (the list), being easier to scan, and (two columns on the screen is) easier

to read. Some said that by being able to see all the options on the screen together, it seemed easier to answer the question. One person (who had the one-column version) said that when he began scrolling, he was thinking, "How long is this list going to be?"

Sponsor Response: A two-column format for the occupation and education items will likely be used in the final Web-based survey.

3.6.2 Medium-Priority Usability Issues

Medium-priority issues typically affect efficiency by extending the amount of time users need to complete tasks. They also can lead to increased frustration and lower satisfaction.

3.6.2.1 The error messages are not prominent.

The error messages are displayed above the main content of the page (i.e., the questions and responses). Several participants did not notice the error message when it appeared. Instead, likely due to the lack of feedback from the display, they were confused and believed the question was being repeated or a very similar question was being asked. They often pressed the 'Next' button several times before realizing an error had appeared. Figure 6 displays three screen shots of items where participants got an error message, with a heat map overlay.
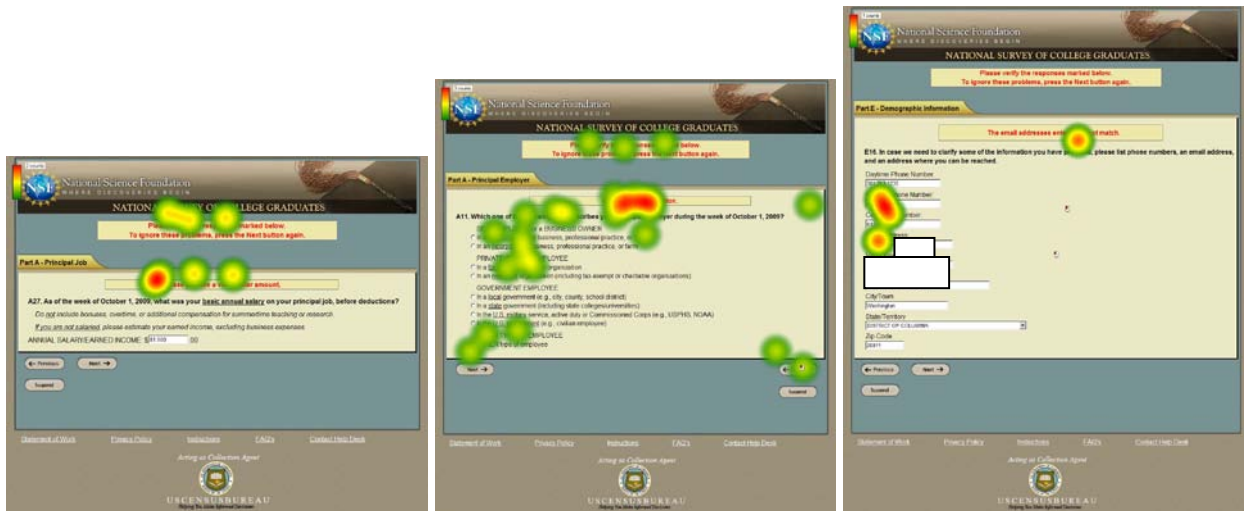
Figure 6. Screen shots of items where participants received an error message. The eye-tracking heat maps demonstrate that participants rarely looked at the uppermost error message, if at all. Note: personal information has been blocked from view in the rightmost image.

It is best to provide the error message in a location that is close to the response that needs to be corrected. Providing the error message within the area where the major content is displayed would be more usable, as participants are likely ignoring content above and below the yellow box, where all of their survey entries are made and where their attention is most often directed.
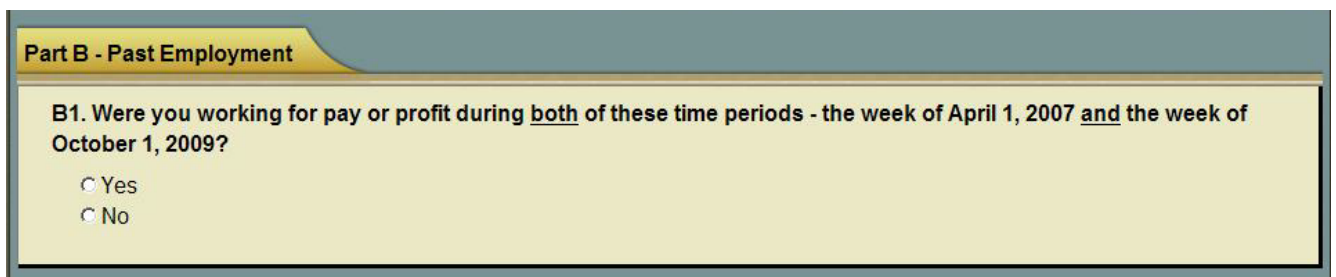
3.6.2.2 Participants do not understand the function of the 'Suspend' button.

None of the participants used the 'Suspend' button during the usability session. When asked during debriefing about what its function might be, many participants were unsure what the 'Suspend' button was for. They often had the general impression that it would pause the survey

so it could be completed at a later time, but they said that did not know how this functionality would be implemented or when they would ever use it. The term 'Save and Exit' used in Iteration 1 (Romano & Chen, 2010) was much clearer to participants. In Iteration 1, participants expected to be logged out of the survey and that their responses would be saved so they could return to complete the survey at a later time.

3.6.2.3 Participants want more information about why certain questions were being asked.

As they worked through the survey, several participants wanted to know why certain questions were being asked. In particular, when items referred to specific dates or weeks of the year, they expressed frustration and surprise because they could not understand why the questions were so specific and because they had difficulty pinpointing what they were doing during that timeframe. One participant said, "Come on, you want to go all the way back to 2007 *and* the week of October 1, 2009? Who's going to know that? The *week* and for it to be that specific...the week of April 1st, 2007." Another participant questioned why the survey was "pinpointing … a specific week instead of a specific month." See Figure 7 for a screenshot of a question that elicited many comments from participants. Of the participants that were asked during debriefing if they would like a short explanation about why certain ranges, dates, or questions were being asked, all said they would. Some participants said they would like information icons throughout the survey that they could click on which would explain the purpose of each question.

Part B - Past Employment

B1. Were you working for pay or profit during both of these time periods - the week of April 1, 2007 and the week of October 1, 2009?

○ Yes
○ No

Figure 7. Participants expressed frustration with the very specific dates and time periods used in Item B1.

3.6.2.4 The instructions for the salary question are unclear.

Several participants mentioned that they were unsure if the salary question (A27, shown in Figure 8) was asking for an exact figure or an estimate. One participant said that she could not remember her salary and questioned, "What should I do? Should I just make a guess?" When uncertain, a few participants said they would pull up appropriate paperwork to find their exact salary if they were at home. Other participants said they wondered how the information was being used and if an exact figure would lead to better data. A few participants mentioned that the question was too personal, although all but one provided a response.

One participant used a period in place of a comma (as shown in Figure 8) as a thousands separator. This is not an uncommon practice but was interpreted by the survey as an invalid dollar amount.

Figure 8. Item A27 does not specify whether income needs to be an exact figure or an estimate.

3.6.3 Low-Priority Usability Issues

Participants can typically recover from low-priority usability issues fairly quickly, but they may become frustrated. Addressing these issues will lead to more satisfied users and a more polished product.

3.6.3.1 Participants interpreted the disability question response options differently.

Users interpreted the scale on the disability question (shown in Figure 9) differently. One participant marked his usual degree of difficulty with seeing as "severe" although he did not appear to be struggling to read words on the Web-based survey. This may be because scale markers such as "slight" and "severe" do not make sense to users and/or are being interpreted differently. Other participants said the scale was confusing, particularly the end-points "none" and "unable to do."

Figure 9. Screen shots of the disability question. Participants had difficulty interpreting the response scales.

Phrases such as "near perfect vision or perfect vision" and "legally blind" would help users determine which response option is best for them. Anchors with more descriptive phrases would also be more usable. For example, in row 1, "no difficulty seeing" would be more informative

than "none," and "blind" would be more informative than "unable to do." Cognitive testing would enable us to understand how typical users interpret the question and responses.

3.6.3.2 There were erroneous codes in many hover tool tips.

Many of the entry fields show extraneous coding in the tool tips. For example, text fields where participants enter in demographic information such as in E16 and E17, all have "<br />" included in the tool tip. In addition, the Next button has the tool tip "submit form" which is erroneous.

3.6.3.3 The large clickable area for the Census Bureau logo in the footer is too large.

A large portion of the area surrounding the Census Bureau logo in the footer is hyperlinked as shown in Figure 10. One participant accidentally clicked on this area which directed him to the Census Bureau homepage. He seemed briefly surprised and then used the back button in the browser to navigate back to the survey.
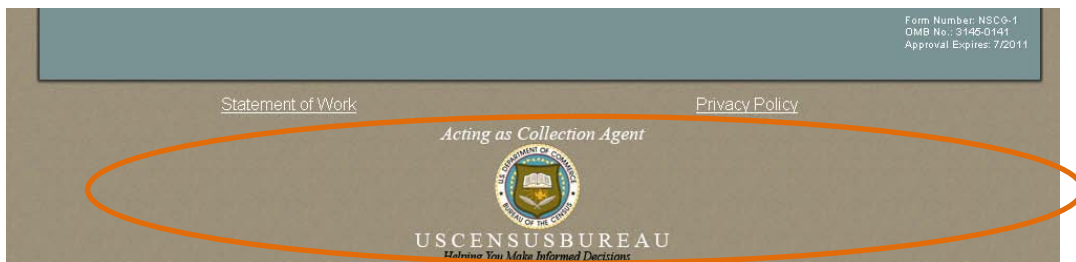
Figure 10. One participant accidentally clicked on the Census Bureau link which has a large clickable area.

4.0 Comparison to Iteration 1 Testing

4.1. Participant Accuracy

None of the participants from either round of testing had difficulty progressing through the survey.

4.2. Participant Efficiency

A re-occurring issue in both rounds of testing is that participants had trouble logging into the survey and navigating to the home page. The numbers of attempts needed to type in the URL and log-in information were similar for both Iterations. See Table 6.

Table 6. Mean Number of Attempts at Entering URL and Log-in Information by Iteration

	Mean number of times to enter text	
Iteration	**URL**	**Log-in information**
1	1.57	1.88
2	1.56	1.60

4.3 Participant Satisfaction

Table 7 displays the mean satisfaction scores for Iteration 1 (Romano & Chen, 2010) and for the NP2 version in Iteration 2 (only the NP2 version was compared to Iteration 1, since it was the same format at the Iteration 1 version).

Table 7. Mean Satisfaction Ratings, Across Participants: Iteration 1 vs. NP2 in Iteration 2

	Mean rating	
	Iteration 1	Iteration 2
Overall reaction to site: terrible - wonderful	6.25	7.00
Screen layouts: confusing - clear	6.38	8.00
Use of terminology throughout site: inconsistent - consistent	6.88	7.38
Information displayed on the screens: inadequate - adequate	6.50	7.13
Questions displayed on the screens: inadequate – adequate	6.00	7.50
Arrangement of information on the screens: illogical – logical	4.57	6.75
Questions can be answered in a straight-forward manner: never – always	6.00	6.88
Organization of question, instruction & response categories: confusing – clear	6.00	6.88
Forward navigation: impossible – easy	6.38	7.88
Overall experience of completing the survey: difficult - easy	7.25	7.50
Mean satisfaction rating, across items, across participants	**6.39**	**7.29**

Participants were more satisfied with the Web-based (NP2) survey in Iteration 2 than in Iteration 1, $t(165) = 2.86$, $p < 0.005$. The average satisfaction score in Iteration 1 was 6.39 out of 9, and in Iteration 2, it was 7.29. Notably, the item "Arrangement of information on the screens: illogical – logical" was rated much higher in Iteration 2. It is unclear why users in Iteration 2 rated their satisfaction higher than the users in Iteration 1.

5.0 Conclusions and Future Directions

Overall, users were able to successfully use the NSCG by completing 100% of the answers on the survey. Minor usability issues have been identified, and the DSD team has made plans to modify the survey. These changes should alleviate many of the important issues and lead to increased usability.

Participants in this usability study were diverse in age, education, and comfort with using Web sites and the Internet. It is probable that individuals who were more comfortable with the Internet and who had higher levels of education were likely to understand the online survey better than those who were at a novice level, but we did not statistically test for this. Future testing should include usability testing on a diverse sample of novices and experts as well as accessibility testing for those with disabilities. It is required for government Web sites to be compliant with Section 508 of the Disabilities Act so that users with disabilities can access public information.

Future research should include another round of usability testing to evaluate whether the changes to the user interface supports users in completing the NSCG online survey. Iterative testing is ideal, in that we test the changes that have been made to see if they lead to improvements in usability and, equally important, if they introduce new issues (Bailey, 1993; Nielsen, 1993; Romano, Olmsted-Hawala, Chen, & Murphy, in prep).

6.0 References

Bailey, G. (1993). Iterative methodology and designer training in human-computer interface design. *Proceedings from ACM CHI 1993 Conference*, 198-205.

Chin, J. P., Diehl, V.A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings from CHI 1988 Conference*, 213-218.

Jans, M. E., Romano, J. C., Ashenfelter, K. T., & Krosnick, J. A. (in prep). Measuring satisfaction of Web sites and Web surveys in the lab: The effects of mode and location on ratings of user satsifaction.

Nielsen, J. (1993). Iterative user interface design. *IEEE Computer, 26,* 32-31.

Poole, A. & Ball, L. J. (2005). Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects. In Ghaoui, Claude (Ed.). *Encyclopedia of Human Computer Interaction*, pp. 211-219. Hershey, PA: Idea Group.

Romano, J. C. & Chen, J. M. (2010). A usability evaluation of the online National Survey for College Graduates (NSCG). U.S. Census Bureau. http://www.census.gov/srd/papers/pdf/ssm2010-05.pdf.

Romano, J. C., Olmsted-Hawala, E. L., Chen, J. M., & Murphy, E. D. (in prep). The challenges and benefits of conducting iterative usability testing on a Web site.

7.0 Acknowledgements

Appendix A. Participant Demographics by Group

Table 8. Participant Demographics by Version

Gender	N	Age	N	Education	N	Race	N
Version NP1							
Male	5	≤ 30	3	Bachelor's	7	African American	4
Female	3	31-45	3	Master's	1	Asian	1
		46-60	1			White	3
		> 60	1				
Mean across Version NP1 participants		**41.75 years***					
Version NP2							
Male	3	≤ 30	3	Bachelor's	6	African American	5
Female	5	31-45	0	Master's	1	American Indian or Alaska Native***	1
		46-60	5	Ph.D.	1	White***	3
		> 60	0				
Mean across Version NP2 participants		**44.38 years***					
Version PN1							
Male	3	≤ 30	0	Bachelor's	4	African American	2
Female	3	31-45	2	Master's	2	White	4
		46-60	2				
		> 60	2				
Mean across Version PN1 participants		**50.83 years***					
Version PN2							
Male	2	≤ 30	2	Bachelor's	4	African American**	4
Female	4	31-45	2	Master's	2	Asian**	1
		46-60	2	Ph.D.	2	White	4
		> 60	2				
Mean across Version PN2 participants		**47.75 years***					
Mean across all participants		**45.87 years***					

*The mean age was calculated from the exact values for each participant. The exact self-reported values were placed in ranges in the table to help the reader get an overview of the data.
**One participant marked African American and Asian.
***One participant marked White and American Indian or Alaska Native.

Appendix B. Questionnaire on Computer and Internet Experience and Demographics

Check (X) for all that apply.

1. Do you use a computer at home or at work or both?
   ___ Home
   ___ Work
   ___ Somewhere else, such as school, library, etc.

2. If you have a computer at home,
   a. What type of internet connection do you use at home?
      ___ Dial up
      ___ Cable
      ___ DSL
      ___ Wireless (Wi-Fi)
      ___ Other _____
      ___ Don't know

   b. Which browser do you typically use at home? Please indicate the version if you can recall it.
      ___ Firefox _____
      ___ Internet Explorer _____
      ___ Netscape _____
      ___ Google Chrome
      ___ Safari
      ___ Other _____
      ___ Don't know

   c. What operating system does the home computer you typically use run in?
      ___ MAC OS
      ___ Windows 95
      ___ Windows 2000
      ___ Windows XP
      ___ Windows Vista
      ___ Windows 7
      ___ Other _____
      ___ Don't know

3. On average, about how many hours do you spend on the Internet per day **total**?
___ 0 hours   ___ 1-3 hours   ___ 4-6 hours   ___ 7or more hours

4. On average, about how many hours do you use the Internet per week **total**?
___ 0 hours   ___ 1-3 hours   ___ 4-6 hours   ___ 7or more hours

5. Which of the following activities do you use the Internet for **more often?**

_____ Searching / Surfing the web **or**   _____ Answering / Sending e-mail

6. Have you ever filled out a survey on the Internet?
   ☐ Yes   ☐ No

a. If yes, about how many surveys do you think you have filled out on the Internet?_____
b. If yes, have you filled out a survey on the Internet in the last two months?
   ☐ Yes   ☐ No

7. How often do you use different media (i.e., photos, audio, video streams) on the Internet?
_____

8. Please rate your overall experience with the following:
    Circle one number.

	No experience							**Very experienced**	
Computers	1	2	3	4	5	6	7	8	9
Internet	1	2	4	5	5	6	7	8	9

9. What computer applications do you use?
    Check (X) for all that apply.
    ___ E-mail
    ___ Internet
    ___ Word processing (MS-Word, WordPerfect, etc.)
    ___ Spreadsheets (Excel, Lotus, Quattro, etc.)
    ___ Databases (MS-Access, etc.)
    ___ Accounting or tax software
    ___ Engineering, scientific, or statistical software
    ___ Other applications, please specify_____

Please circle one number for each question below.	**Not at all Comfortable**				**Very Comfortable**
10. How *comfortable* are you in learning to navigate new Web sites?	1	2	3	4	5
11. Computer windows can be minimized, resized, and scrolled through. How *comfortable* are you in manipulating a window?	1	2	3	4	5
12. How *comfortable* are you using and navigating through the Internet?	1	2	3	4	5
13. How *comfortable* are you in providing personal information through the Internet?	1	2	3	4	5

	Never				**Very Often**
14. How *often* do you work with any type of data using a computer?	1	2	3	4	5
15. How *often* do you perform complex analyses of data through a computer?	1	2	3	4	5
16. How *often* do you use the Internet or Web sites to find information? (printed reports, news articles, data tables, blogs, etc.)	1	2	3	4	5
17. How *often* do you use the Internet for personal reasons (e.g. Facebook, blog, etc.)?	1	2	3	4	5
18. How *often* do you use the Internet for business (e.g. online banking, etc.)?	1	2	3	4	5

	Not at all Familiar				Very Familiar
19. How *familiar* are you with the Census Web site (location, tools, terms, data, etc)?	1	2	3	4	5
20. How *familiar* are you with the National Survey of College Graduates survey, National Survey of Recent College Graduates or Survey of Doctoral Recipients?	1	2	3	4	5

21. What is your date of birth?   _____
                                                          month   year

22. What is the highest grade of school you have completed, or the highest degree you have received?
     a) [ ] Completed ninth grade or below
     b) [ ] Some high school, but no diploma
     c) [ ] Completed high school with diploma or received a GED
     d) [ ] Vocational training beyond high school
     e) [ ] Some college credit
     f) [ ] Associates degree (AA/AS)
     g) [ ] Bachelor's Degree (BA/BS)
     h) [ ] Master's degree (MA/MS)
     i) [ ] Professional degree
     j) [ ] Doctoral degree

     For options D through J above, indicate area of study: _____

23. What is your gender?

_____ Male   _____ Female

24. Do you consider yourself to be of Hispanic, Latino, or Spanish origin?
(Optional. We ask this question to ensure a diverse sample of people is in each study.)

_____ Yes   _____ No

25. What is your race? Choose **one or more** races.
 (Optional. We ask this question to ensure a diverse sample of people is in each study.)

_____ White
_____ Black or African American
_____ Asian
_____ Native Hawaiian or Other Pacific Islander
_____ American Indian or Alaska Native

Table 9. Participants' Self-Reported Computer and Internet Experience

Participant	Hours per day online	Filled out online survey	Number of online surveys taken	Overall experience with computers	Overall experience with Internet	Comfort in learning to navigate new Web sites	Comfort in manipulating a window	Comfort in using and navigating the Internet	Comfort in providing personal info online	How often working with data on a computer	How often working with analyses of data with a computer	How often using the Internet or Web sites to find info	How familiar with the Census Web site (location, tools, data, etc)	How familiar with NSCG, NSRCG, or SDR
				Scale: 1 (no experience) – 9 (very experienced)		Scale: 1 (not comfortable) – 5 (comfortable)				Scale: 1 (never) – 5 (very often)			Scale: 1 (not familiar) – 5 (very familiar)	
NP1- 1	7+	Yes	50+	7	7	4	5	4	2	4	2	4	2	1
NP1- 2	1-3	Yes	20	9	9	5	5	5	5	5	3	5	2	1
NP1- 3	7+	Yes	10	6	7	4	4	4	2	5	2	5	1	1
NP1- 4	1-3	Yes	3	5	8	3	4	4	3	2	1	3	3	1
NP1- 5	4-6	Yes	5	8	8	5	5	5	4	3	2	4	1	1
NP1- 6	4-6	No	-	5	5	4	2	3	1	2	1	3	1	1
NP1- 7	1-3	Yes	10	3	5	3	3	4	3	3	2	4	1	1
NP1- 8	4-6	Yes	10	6	6	4	4	4	4	3	2	4	1	1
NP2- 1	7+	Yes	10+	9	9	5	5	5	2	5	3	5	1	1
NP2- 2	1-3	Yes	50	5	6	4	5	5	4	4	2	2	1	1
NP2- 3	4-6	No	-	4	6	3	3	4	4	3	1	3	3	2
NP2- 4	4-6	Yes	4	9	9	5	5	5	3	2	2	5	1	1
NP2- 5	4-6	Yes	40	1	9	5	5	5	5	4	4	5	4	1
NP2- 6	4-6	Yes	10	7	7	5	5	5	3	5	5	5	4	3
NP2- 7	4-6	Yes	3	6	6	4	3	4	3	4	1	3	2	1
NP2- 8	7+	Yes	10	7	7	4	4	4	3	5	4	5	2	1
PN1- 1	1-3	No		6	6	4	3	3	2	4	3	3	1	1
PN1- 2	4-6	Yes	10	9	9	5	5	5	5	5	3	5	2	1
PN1- 3	1-3	Yes	2	9	9	5	5	5	3	3	3	5	3	1
PN1- 4	1-3	Yes	25-30	7	7	4	3	4	4	2	1	5	1	1
PN1- 5	1-3	Yes	12	5	5	3	4	4	2	5	4	5	2	1
PN1- 6	1-3	No		7	7	5	5	5	1	2	2	4	2	1
PN2- 1	7+	Yes	20	9	9	5	5	5	2	5	3	5	3	1
PN2- 2	1-3	Yes	4	9	9	5	5	5	3	5	3	5	5	1
PN2- 3	1-3	Yes	30	8	8	5	5	5	4	5	5	5	3	3
PN2- 4	4-6	Yes	5	7	7	4	4	4	2	3	2	4	2	1
PN2- 5	4-6	Yes	15-20	7	7	4	5	5	3	5	4	5	1	1
PN2- 6	4-6	Yes	5	7	7	4	4	4	4	5	3	5	1	1
PN2- 7	4-6	Yes	2	5	6	5	5	4	2	5	3	5	3	1
PN2- 8	1-3	No	-	8	9	5	5	5	1	4	1	5	1	1
Mean				**6.71**	**7.35**	**4.35**	**4.35**	**4.45**	**2.94**	**3.90**	**2.55**	**4.39**	**2.00**	**1.19**

Appendix C. Researcher Script for Usability Session

Thank you for your time today. My name is (Test Administrator). I work here in the U.S. Census Bureau Usability Lab, and I will be working with you today. In this lab, we evaluate how easy or difficult Census products are to use. We bring in people like you who are potential users of our products to try them out while there is still time to make changes to them. What works well, we keep. When potential users such as you have difficulty with something, we have an opportunity to fix it.

Today, we will be evaluating the National Survey of College Graduates by having you work on the survey. There are two parts to our session. First, you will complete the Web survey. Then, at the end of the session, you will fill out a questionnaire about your experience during the session. The entire session should last about an hour.

Before we start, there is a form I would like you to read and sign. It explains the purpose of today's session and your rights as a participant. It also informs you that we would like to videotape the session to get an accurate record of your feedback. Only those of us connected with the project will review the tape and it will be used solely for research purposes. Your name will not be associated with the tape or any of the other data collected during the session.

[Hand consent form; give time to read and sign; sign own name and date.]

Thank you.

Before we start, I want to tell you that you can't make a mistake or do anything wrong here. Difficulties you may run into reflect the design of the survey, not your skills or abilities. This product is intended for people like you. Where it works well, that's great. If you have a problem using parts of it, that is also great, because you will help us to identify these places. We are going to use your comments and data as well as comments and data from the other participants to give feedback to the developers of the Web survey. Your comments and thoughts will help the developers make changes to improve the survey. I did not create the survey, so please do not feel like you have to hold back on your thoughts to be polite. We are not evaluating you or your skills, but rather you are helping us see how well the Web survey works. Please share both your positive and negative reactions to the survey.

Your comments are very important to us. I'd like you to tell me your impressions and thoughts as you work through the survey. So give me your open impressions, both good and bad of what you see and what you experience.

While you are working, I would like you to think aloud. In other words, I'd like you to tell me what you are thinking, describe the steps you are taking, what you are expecting to see, why you are doing what you are doing, what you are going to do, and why. Tell me why you clicked on a link or where you expect the link to take you. Tell me if you are looking for something and what it is and whether you can find it or not.

Do you have any questions about the "think-aloud" process?

Tell Participant about interruption that will take place during the survey.

Tell Participant about the letter that they would receive in the mail- when you ask them to, they are to read it and proceed with the survey as they would if they got the letter in the mail.

Ok, now we are ready to begin. I am going to go around to the other room and do a sound check. I will instruct you when to read the instructions on this paper and when to begin. Also, here is the questionnaire you will complete at the very end. I will tell you when to complete this.

[Hand computer experience form. Set the task questions and questionnaire by participant.]

I am going to leave, but we will still be able to communicate through a series of microphones and speakers. Do you have any questions?

[Go to control room. Do sound check. Start video recording.]

Now we are ready to begin. Please begin by reading the instructions out loud. As you work, remember to talk to me about what you are thinking and feeling.

Appendix D. Consent Form

Consent Form
For Individual Participants
Usability Testing of the NSCG Web Survey

Each year the Census Bureau conducts many different usability evaluations. For example, the Census Bureau routinely tests the wording, layout and behavior of products, such as Web sites and online surveys and questionnaires in order to obtain the best information possible.

You have volunteered to take part in a study to improve the usability of the NSCG Web Survey. In order to have a complete record of your comments, your usability session will be videotaped. We plan to use the tapes to improve the design of the product. Only staff directly involved in the research project will have access to the tapes. Your participation is voluntary and your answers will remain strictly confidential.

This usability study is being conducted under the authority of Title 13 USC. The OMB control number for this study is 0607-0725. This valid approval number legally certifies this information collection.

I have volunteered to participate in this Census Bureau usability study, and I give permission for my tapes to be used for the purposes stated above.

Researcher's Name: _____

Researcher's Signature: _____ Date: _____

Participant's Name: _____

Participant's Signature: _____ Date: _____

Appendix E. Log-In Letter

Dear Respondent,

As a college graduate in a science or engineering field, you have been selected to participate in the National Survey of College Graduates. The National Science Foundation, together with the US Census Bureau, is pleased to offer a new and convenient way to respond. Responding by Web saves you time and is easy to use.

To complete the survey online, follow the three easy steps listed below:

1. Go to **http://idc4.ssd.census.gov:4321/nscg** (This is a secure site).
2. On the opening screen, enter the following Username and Password.

Username: NSCG6745

Password: $eaFPq93YjDr

3. Once you log in, follow the instructions on the screen to complete the questionnaire.

If you have any questions about the study, please contact the Census Bureau between 9 a.m. and 5 p.m. EST at 301-763-3856 or send an e-mail to nscg@census.gov.

Your participation in this important research effort is greatly appreciated.

Thank you.

Appendix F. Satisfaction Questionnaire

Please circle the numbers that most appropriately reflect your impressions about using this Web survey.

1. Overall reaction to the Web site:

terrible   wonderful
1 2 3 4 5 6 7 8 9 not applicable

2. Screen layouts:

confusing   clear
1 2 3 4 5 6 7 8 9 not applicable

3. Use of terminology throughout the Web survey:

inconsistent   consistent
1 2 3 4 5 6 7 8 9 not applicable

4. Instructions displayed on the screens:

inadequate   adequate
1 2 3 4 5 6 7 8 9 not applicable

5. Questions displayed on the screens:

inadequate   adequate
1 2 3 4 5 6 7 8 9 not applicable

6. Arrangement of necessary information (FAQ, Help, etc.) on the screen:

illogical   logical
1 2 3 4 5 6 7 8 9 not applicable

7. Questions can be answered in a straight-forward manner:

never   always
1 2 3 4 5 6 7 8 9 not applicable

8. Organization of question, instruction and response categories in the survey:

confusing   clear
1 2 3 4 5 6 7 8 9 not applicable

9. Forward navigation:

impossible   easy
1 2 3 4 5 6 7 8 9 not applicable

10. Suspending and resuming the survey:

impossible   easy
1 2 3 4 5 6 7 8 9 not applicable

11. Overall experience of completing the survey:

difficult   easy
1 2 3 4 5 6 7 8 9 not applicable

Additional Comments

Table 10. Mean Satisfaction Questionnaire Responses by Version (1 = low, 9 = high)

					Satisfaction Questionnaire Items							
Participant	Overall reaction to site: terrible - wonderful	Screen layouts: confusing - clear	Use of terminology throughout site: inconsistent - consistent	Information displayed on the screens: inadequate - adequate	Questions displayed on the screens: inadequate - adequate	Arrangement of information on the screens: illogical - logical	Questions can be answered in a straightforward manner: never - always	Organization of question, instruction & response categories: confusing - clear	Forward navigation: impossible - easy	Suspending and resuming the survey: impossible - easy	Overall experience of completing the survey: difficult - easy	Mean rating by participant
Version NP1												
1	6	8	8	8	8	4	7	9	9	*	9	**7.60**
2	9	9	9	9	9	9	9	9	9	*	9	**9.00**
3	5	8	7	6	8	4	2	4	5	4	8	**5.55**
4	2	4	3	6	6	7	4	4	9	*	3	**4.80**
5	9	9	8	9	8	9	8	8	9	*	9	**8.60**
6	5	9	7	9	9	9	8	9	9	*	8	**8.20**
7	7	8	7	7	8	8	7	7	8	*	8	**7.50**
8	7	8	8	8	8	8	7	8	8	*	8	**7.80**
	6.25	**7.88**	**7.13**	**7.75**	**8.00**	**7.25**	**6.50**	**7.25**	**8.25**	**4.00**	**7.75**	**7.36**
Version NP2												
1	7	7	9	6	6	7	4	7	9	*	8	**7.00**
2	8	9	8	9	9	8	9	9	9	*	9	**8.70**
3	6	7	6	6	8	4	7	6	6	*	6	**6.20**
4	6	8	2	5	5	3	6	3	5	*	5	**4.80**
5	7	6	9	7	7	8	8	7	8	*	7	**7.40**
6	8	9	9	9	9	9	6	8	9	*	9	**8.50**
7	7	9	8	8	8	8	8	8	9	*	8	**8.10**
8	7	9	8	7	8	7	7	7	8	*	8	**7.60**
	7.00	**8.00**	**7.38**	**7.13**	**7.50**	**6.75**	**6.88**	**6.88**	**7.88**	*****	**7.50**	**7.29**

Version PN1												
1	8	9	8	8	8	8	6	6	7	6	7	**7.00**
2	8	9	9	9	9	8	9	9	9	*	9	**8.80**
3	8	9	6	8	5	N/A	5	8	9	*	8	**7.33**
4	8	7	9	8	8	N/A	7	9	9	*	8	**8.11**
5	9	9	9	9	9	9	9	8	9	*	9	**8.90**
6	8	9	8	8	8	8	7	8	9	*	9	**8.20**
	8.17	**8.67**	**8.17**	**8.33**	**7.83**	**8.25**	**7.17**	**8.00**	**8.67**	**6.00**	**8.33**	**8.12**
Version PN2												
1	8	7	8	8	6	5	7	7	8	8	7	**7.18**
2	8	7	5	8	9	9	7	7	9	9	8	**7.82**
3	7	9	8	9	9	9	8	7	9	9	9	**8.45**
4	4	7	7	7	7	5	7	6	8	*	6	**6.40**
5	7	9	7	7	7	N/A	7	*	9	8	9	**7.78**
6	8	8	9	9	8	9	9	9	9	*	9	**8.70**
7	9	9	8	8	7	9	7	8	9	*	8	**8.20**
8	9	9	9	9	9	9	8	9	9	9	9	**8.91**
	7.50	**8.13**	**7.63**	**8.13**	**7.75**	**7.86**	**7.50**	**7.57**	**8.75**	**8.60**	**8.13**	**7.94**
Mean rating by question across all participants	**7.17**	**8.13**	**7.53**	**7.80**	**7.77**	**7.41**	**7.00**	**7.38**	**8.37**	**7.57**	**7.89**	**7.65**

* = Question removed from questionnaire

Blank = Participant left item unanswered

N/A = Participant chose N/A option on questionnaire

Appendix G. Debriefing Questionnaire

1. Can you walk me through your thinking on why you marked (a particular QUIS item) especially low/high? (Do this for several low/high QUIS ratings).

2. What was your first impression of the Web survey Web site?

3. What was your overall experience from logging into the Web survey to the first survey question?

4. What do you think of the basic screen layout?   a. Overall?   b. Colors?
        c. Links and information around the center pane?
        d. Context of the information on the homepage?
        e. Other?

5. What do you think of the navigational methods?   a. Previous and Next buttons?
        b. Saving and Exit button?

6. What was your experience of logging back in after suspending the survey?

7. What do you think of the questions asking you to select a job category?
      a. Was it easy for you to find and select the appropriate category?
      b. Do you feel confident that you selected the best category?
      c. Would you prefer the answers to be displayed in 2 columns or 1 column?

8. What do you think of the questions asking you to select an education category?
      a. Was it easy for you to find and select the appropriate category?
      b. Do you feel confident that you selected the best category?

9. How easy or difficult was it to complete the survey in general?

10. What did you think of the follow-up questions that prompt you for an answer when you did not provide an answer the first time? (if necessary)

11. *If participant clicked on any of the links in the footer, ask:* Why did you click on X link?
   If participant did not click on any of the links in the footer, ask: What do you think of these links?

12. What did you like best about the Web survey?

13. What did you like least about the Web survey?

14. Is there anything that you feel should be changed?

15. Is there anything that you feel should stay the same?

16. How easy/difficult do you feel it was to complete the questions? What made a question easy/difficult?

17. Was there any information you needed that you could not find or was not provided on the screen?

18. Is there anything you would like to mention that we haven't talked about?

Appendix H. Gaze plots for each participant on the log-in screen demonstrating that most people did not read the text on the log-in screen.

Appendix I. Gaze plots for each participant on the Introduction screen, demonstrating that most people read the text on the Introduction screen, which is the first screen respondents see following a successful log in.

Appendix J. Gaze plots for each participant for the one-column version of the occupation code selection item.

Appendix K. Gaze plots for each participant for the two-column version of the occupation code selection item.