Proceedings of Statistics Canada Symposium 2001
Achieving Data Quality in a Statistical Agency: A Methodological Perspective

USABILITY TESTING AND COGNITIVE INTERVIEWING TO SUPPORT ECONOMIC FORMS DEVELOPMENT FOR THE 2002 U.S. ECONOMIC CENSUS

Amy E. Anderson, Elizabeth Nichols, Kimberly Pressley[1]

ABSTRACT

Increasing demand for reporting electronically in establishment surveys has placed additional emphasis on incorporating usability into electronic forms. We are just beginning to understand the implications surrounding electronic forms design. Cognitive interviewing and usability testing are analogous in that both types of testing have similar goals: to build an end instrument, whether it is paper or electronic, which reduces both respondent burden and measurement error. Cognitive testing has greatly influenced paper forms design and can also be applied towards the development of electronic forms. Usability testing expands on existing cognitive testing methodology to include examination of the interaction between the respondent and the electronic form.

The upcoming U.S. 2002 Economic Census will offer businesses the ability to report using electronic forms. The U.S. Census Bureau is creating an electronic forms style guide outlining the design standards to use in electronic form creation. The design standards outlined in the style guide are based on usability principles, Graphical User Interface design standards, and results from usability and cognitive testing. This paper highlights the major electronic forms design issues raised during the preparation of the style guide and describes how usability testing and cognitive interviewing resolved these issues.

KEY WORDS: Establishment surveys, Usability testing, Cognitive interviewing

1. INTRODUCTION

Every five years, in years ending in 2 and 7, the U.S. Census Bureau uses the economic census to capture in-depth information about the economy. The information collected is necessary for a variety of data users such as federal and state governments, researchers, academics, and the private sector. The data collected are used by economists to calculate key economic indices, is referenced by legislators in the creation and revision of laws, and is examined by companies before making important business decisions. These and a myriad of other important uses highlight the importance of collecting quality data from businesses. The 2002 Economic Census will include over 600 forms covering retail, wholesale, manufacturing, mining, services, and construction industries.

In previous censuses, data collection has primarily been mailed self-administered paper forms. The 1997 Economic Census introduced the Computerized Self Administered Questionnaire (CSAQ) as a supplemental means of data collection for selected companies reporting on retail industry forms. The 2002 Economic Census is planning to offer all companies electronic reporting in addition to the traditional paper forms. A goal in designing both paper and electronic forms is to ensure consistency. To accomplish this across all electronic forms, a team of Census Bureau employees was brought together to create a style guide. Style guides provide rules for developers during electronic form design. The style guide we

---

[1] Amy E. Anderson, Elizabeth Nichols, Kimberly Pressley U.S. Census Bureau, Washington D.C. 20233

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

developed draws from researched usability principles, design standards, and best practices from other business electronic forms that have been created within the Census Bureau, and usability tests.

This paper describes two different rounds of usability testing that were completed to support the style guide. The first round of testing focused on proposed wording of edit messages. The second round of testing was a more comprehensive test of layout guidelines.

2. EDIT MESSAGE TESTING

2.1 Goals

The first round of usability testing focused on proposed wording of edit messages displayed to respondents when data entered failed edit tests built into the instrument. The style guide team identified the different types of edits possible in the electronic form and proposed wording for those edits. The team wanted to ensure that all respondents interpreted the messages similarly and without major comprehension problems. The major goals in testing were to examine the wording of the edit messages, the style in which the messages were presented, and the timing in which the edits would first appear to the respondents.

2.2 Methodology

Ten concurrent think-aloud laboratory interviews were conducted to test respondents' understanding of proposed edit wording for a fictitious on-line mortgage application. Staff from the economic area of the Census Bureau was recruited as participants. To avoid respondent fatigue, interviews were limited in length to one hour. Interviews were conducted in the Usability Laboratory at the Census Bureau, and, with the exception of one interview, all interviews were tape-recorded.

The on-line mortgage application was chosen for testing because it was felt that Census Bureau staff could relate to that topic. In addition, our proposed edits fit the context of the mortgage application, and adjustments to the wording of the edits to fit this application was not necessary.

Participants were given a paper prototype of an on-line mortgage application. Paper prototypes are typically mock designs, often created in a word processing application, which mimics the look of the final computer application screens. The mock on-line mortgage application was pre-filled with data and participants were instructed to read through the form thoroughly to become familiar with it and the information requested in the application.

Three of the edits presented to the respondents as they read through the application were preventative edits. They were displayed to the respondent at the time that they occurred. For example, suppose the respondent enters in their telephone number as a string of numbers (e.g., 7034584589). The computer application may be expecting dashes to be placed between the numbers (e.g., 703-458-4589). The computer would then give the respondent an edit message at the time they were responding to the question.

Once the participants finished reading through the application they were presented with a list of twelve edit messages similar to ones that they would have received had they submitted the application through the Internet. Participants were then asked to read through each of the edit messages one at a time and were asked one or more of the following questions regarding each message:

- What does that message mean to you?
- What do you feel you need to do? [go ahead and do that]
- Why are you getting this message?
- What would you do to resolve this message?

The style of the wording within the edits was also tested. Inclusion of the question number within the edit wording was varied across the edits. Capitalization of the variables, similar to that which would be loaded in a metadata database, was also varied across messages. There were several different sorts of edits tested: range edits, detail missing edits, inter-item edits, etc.

After going through all of the edit messages, participants were asked four overall questions:
- Do you have any thought about the format and layout we used for the messages?
- Thinking back over all the messages, which messages were the easiest to understand, the most difficult, and why?

2.3 Limitations

There were several limitations in our testing of edit wording. The internal Census Bureau employees used were not representative of the respondents that typically fill out economic census forms. Actual business respondents may have behaved differently when presented with the edit messages and may have interpreted them differently then the Census Bureau employees had. Results may differ for applications other than our fictitious on-line mortgage application. The context in which the edit messages were understood would have been different if tested in conjunction with an economic census form. Reaction and feedback about the messages may also differ. Finally, differences in interpretations could have occurred if the form was presented on a computer screen versus on a piece of paper. However, as testing was primarily focused on the interpretation of the messages themselves, and not so much with the usability of the form, using a paper prototype may have been sufficient.

2.4 Results

A set of general recommendations was compiled based on feedback from respondents about the content and style of the tested edit messages. These recommendations were then adapted for inclusion in the style guide. The following are a selection of the recommendations for edit messages based on testing:

- **Provide a location, description of the problem, and an action to take.** The content of each error message was varied throughout the list of edits. Some error messages were presented without key pieces of information such as which question number(s) the message was referring to and what type of action was needed in order to resolve the error. Respondents reacted to this by spending extra time locating the source of the error and figuring out how to remedy the problem. Subject matter content was often used as a source of additional information for the respondent to use in locating the appropriate question.
- **Beware of mismatches between language used in the messages and questions.** For testing purposes, there were several error messages where the subject-matter content was changed slightly. Typically subject-matter content should match back to the information provided in the actual question. Respondents found this mis-match to be another obstacle in dealing with the error messages.
- **Avoid jargon.** Avoiding jargon is a principle rule for both designing questionnaires as well as designing computer applications. Including jargon in an application or questionnaire can impede respondents from completing their task. When the respondents were presented with jargon, they reacted as expected and wasted time trying to interpret the meaning of the jargon used.
- **Be polite.** The word “please” was used throughout many of the messages tested in an effort to convey a pleasant tone and to reduce any negative associations that could have been made.
- **Be sensitive to the respondent:** There are words that are often used when presenting error messages that can be construed by respondents as judgmental. Such words include “error”, “discrepancy”, and “mistake”. Use of these words when designing the error messages to test was avoided.
- **Beware of offering only one of many possible solutions.** Reminders were used in several of the error messages that instructed respondents to “please check for typing mistakes.” Although these messages were meant to be helpful for the respondents, the reminders turned out to be distractions. Many respondents dwelled too long on the possibility of their error being a typing mistake.
- **Capitalize variable names.** There were several variable names referring to subject-matter content that were included in the error messages. Respondents used these variables to help

them understand the meaning of the error message as well as using them as an indicator for where the error may be located when a question number was missing. The respondents noted that it would make their task easier if the variable names within the error messages stood out more.

- **Keep messages brief.** Respondents preferred shorter messages that conveyed the relevant information quickly as opposed to messages that were wordier and provided more detail. But as stated above, respondents need key pieces of information as well. So it is recommended that edit message wording be as succinct as possible.
- **Use good grammar.** Often when variables are just placed in sentences, the overall grammar of the sentence becomes very awkward. Sometimes this can be remedied by the careful creation of variable names.
- **Present edit messages as soon as possible.** The bulk of the edits were presented after submission of the application because it was felt that being presented with edits continuously would bother respondents. During testing respondents felt the opposite. They felt additional burden when presented with a long list of error messages that they then had to read through and fix.
- **Format data-entry fields to prevent errors.** Respondents were quick to note that the preventative edits that were given to them for their social security, date of birth, and zip code could have easily been avoided had the question itself provided the correct format in which to respond.

Table 1 presents some examples of how the recommendations from the testing of edit messages changed the original wording that was tested.

Table 1. Examples of Tested Wording and Recommended Wording

Tested Wording	Recommended Wording
Homeowner's Fee is $1,200. Typical values are between $25 and $300. Please check for typing mistakes. If entries are correct, explain discrepancy in the remarks section.	In Item 1, typical values are between $25 and $300. You reported $1200. If this value is correct, please explain in the remarks section.
Since you reported found a new home in item 15, please complete signed a sales contract in item 16.	Please complete Item 3, since you completed Item 4.
Sum of down payment percentages is 110%. It should be 100%. If values are correct, explain discrepancy in the remarks section.	In Item 7, the values you reported add up to 110%. The total should be 100%.

3. TESTING OF RETAIL BUILDING SUPPLIES PROTOTYPE

3.1 Goals

An electronic prototype of a retail census form was created for the second round of testing for the style guide. The prototype was designed using rules that had been proposed in the style guide. The team was interested in testing the prototype in an effort to identify any problematic layout decisions under consideration. Some of the goals in testing were to look at the layout of the form, it's functionality, and the navigation through the form.

3.2 Methodology

Three concurrent think-aloud usability tests of a protyped 2002 Economic Census electronic form for the retail industry were conducted. The participants recruited for testing were from small multi-unit companies in the Washington, D.C. area. The usability interviews typically lasted one and a half hours.

Recruitment of additional respondents was attempted, but there was a lack of local multi-unit retail building supply companies in the Washington D.C. area for which the electronic form was built. The majority of local companies in the retail sector are single unit companies. Conducting interviews with single location companies likely provides feedback of lower quality since these companies tend to report less frequently via electronic instruments. On the other hand, since all business will be offered the opportunity to report electronically for the 2002 Economic Census, testing with single units likely provided some insights about novice users.

Respondents were presented with the prototype retail census form on a laptop computer that was provided by the researchers. Since a majority of the questions on the form required reference to detailed records, respondents were asked to complete the form to the best of their ability by making estimates. Respondents were also asked to look at the help included in the prototype that was created using an off-the-shelf product called For Help. They were also presented with a paper mock-up of a Survey In-Box. The Survey In-Box is a separate window that lists the forms associated with each establishment within a company. Companies will be able to view their establishments, select which establishment to report for, import their data, print selected forms, and submit their data within the Survey In-Box.

The final topic of discussion during the usability testing was importing. Importing data allows respondents to set up a separate spreadsheet populated with the data requested and bring that data into the form without having to individually go through each individual form and question. Large multi-unit companies have found this to be an ideal and time saving way for reporting data. Due to the small size of the companies tested, participants were not able to comment about importing.

3.3 Limitations

There are several limitations that should be considered when interpreting the results of this testing. The testing was done solely with local companies. These companies might not be necessarily representative of the types of companies in the building materials and supplies dealers industry.

A very small number of companies were recruited for this testing. Other companies may not have had the same problems that these three companies had, or could have uncovered additional problems beyond those revealed by the three companies. The three companies tested however did show consistent patterns which adds merit to our recommendations.

The companies that were tested were small multi-unit companies. The respondents managed the accounting for at most five separate locations for their company. They noted that it would be unlikely that they would chose to report electronically if given the opportunity. They also noted that it would be overly time-consuming for them to set up their data to import into the electronic form. Large companies may characterize and deal with their locations differently than small companies. Large companies would be more apt to report electronically and would find it more time-efficient to import their data. The testing may have involved a more representative sample had large multi-unit companies been recruited.

3.4 Results

Overall, respondents reacted very positively to the electronic prototype of the retail census form. Results based on feedback from respondents and observations were presented to the style guide team and ordered by priority. The following is a summary of selected high priority findings:

- **The functionality of the Survey In-Box was unfamiliar to respondents:** The prototype of the Survey In-Box as presented to respondents did not contain usage instructions. By just looking at the paper prototype, respondents did not immediately understand what to do. It was recommended that instructions be included somewhere within the Survey In-Box window or that a link to instructions be provided.

- **Submitting via the Survey In-Box was not clear.** Whether or not respondents could submit their responses one establishment at a time or all at one time was not clearly stated in the Survey In-Box. Instructions are needed to clarify the process of submitting data.
- **Check boxes did not function as mutually exclusive for "choose-one" entries.** The use of radio buttons was prohibited due to problems when respondents tab within and between questions. Instead, check boxes were incorporated into the form for questions where respondents needed to choose several entries or just one. Check boxes allow respondents to choose several options even when the instructions request a single choice. Respondents were surprised when they could click on more than one response to questions requiring only one. Developers are looking into the possibility of creating an "enhanced check box" that would function similarly to a radio button, but would not move selections when the tab key is used.
- **The definition of the reporting unit wasn't apparent to all respondents.** The reporting unit instruction that appears on the paper form was moved in the electronic form to a Welcome Page. At least one respondent missed this instruction because of its placement on the welcome page. It is recommended that the placement of this important instruction be moved above the first question.
- **Respondents wanted item-specific information a click away.** There are several items on the form that have additional instructions. In paper forms, these instructions are on a separate information sheet that accompanies the paper form. In the electronic form, these instructions are imbedded into a help file. For one of the items on the prototype, a link was given for the item specific instructions right at the item. Respondents preferred accessing item-specific help this way rather than going through several additional actions when accessing help.
- **Respondents may not be sure what to do at the end of the survey form.** The final page of the prototype just had a button which said "Previous." At this point in the form, respondents noted that they weren't sure where to go next to complete remaining forms. They thought that they may have to return to the Survey In-Box in order to do this, but it was not clear how to return to the Survey In-Box. It was recommended that additional buttons be added to the final screen for each survey. One button could take respondents directly to the next form. Another button would return respondents to the Survey In-Box. The need for returning to the Survey In-Box at any point in the survey was also noted, so a recommendation was given for including the "Return to Survey In-Box" button on each page of the survey.
- **Manual summation of figures is tedious for respondents.** Respondents commented that their reporting task would be made much simpler if the computer application would automatically sum figures where appropriate.
- **The two-column layout of the item requesting detailed receipts data or detailed merchandise line could be a problem for respondents.** Two columns are available to respondents for reporting figures. They have the option of reporting in either dollar figures or percentages. Respondents are supposed to just report in one column and not both. They have the option of reporting in either dollar figures or percentages. All of the respondents tested felt that they should be able to report in both columns. The instructions did not clarify the task for respondents. Several suggestions were proposed to fix possible problems with this item. One column could be eliminated and a pull-down added that would require respondents to choose between reporting dollar figures or percentages. The instructions for this item should provide further clarification. Two columns could remain, but once the respondent begins entering information into one column, the other becomes disabled.

4. CONCLUSIONS

In conclusion, the contents of the style guide were improved through usability testing of proposed edit wording and a prototyped census form. These results, along with standard usability principles and knowledge from past usability tests conducted on other economic surveys, were incorporated into the 2002 Economic Census Electronic Forms Style Guide. Although more informed layouts are being designed with guidance from the recommendations from testing. Continued improvements to the style guide could be gained through further testing.