The Effect of Data Collection Software on the Cognitive Survey Response Process

Rebecca L. Morrison, Amy E. Anderson, Charles F. Brady
U.S. Census Bureau

ABSTRACT

The traditional four-step cognitive response process (comprehension, retrieval, judgment, and communication) was expanded for establishment surveys to accommodate organizational-level factors such as multiple respondents, reliance on business records, and competing reporting requirements. Ideally, establishment survey respondents know where to find requested information, and can translate it easily from records to the questionnaire.

Electronic data collection adds complexity to the response process: respondents interact with the question, their records, and also with the electronic instrument. The matter is complicated even further when the data collection software is unfamiliar. Since electronic reporting requires that the data be put into a specific format, additional cognitive burden is expended because of the need to understand the instrument, its navigation, and its requirements.

This paper describes the effects of electronic data collection on the establishment response process, based on research by the U.S. Census Bureau.

Keywords: establishment surveys, usability, electronic reporting, questionnaire design

1. Introduction

The cognitive response process identifies the steps that respondents move through in order to respond to a data request. The survey design community generally understands these steps –comprehension, retrieval, judgment, and communication/reporting – and some surveys have experienced improvements based on knowledge about the cognitive response process. This process was originally applied to household and social surveys, and accommodations for establishment surveys have been made in recent years to that survey setting.

Both the cognitive response process and the establishment response process were developed while electronic reporting was in its infancy. Their focus was on paper survey instruments. However, electronic reporting calls for tasks that are unique to that mode, and which affect the establishment response process model. Respondents go through additional steps that are not found in either the traditional model or the establishment model. Our paper describes findings that were revealed during research conducted as part of the data collection software improvement process undertaken at the U.S. Census Bureau. In addition, we examine the effect of electronic reporting on the establishment response process model.

2. Background

The steps of the cognitive response process, outlined by Tourangeau (1984), are comprehension, retrieval, judgment, and communication/reporting. Each step is described below.

Comprehension refers to the respondent's interpretation and understanding of the question's language, structure, and grammar. In order to answer the question, a respondent must understand what information is being requested. Retrieval is the step where relevant information is obtained, either from records or from memory. The next step, judgment, describes the respondent's evaluation of the completeness or relevance of the data obtained. It is here that approximations are made based on partial or incomplete data. The last step, communication or reporting, deals with mapping the response to the answer space provided and possibly editing the answer.

In the years since Tourangeau's initial foray into the cognitive response process, several other researchers have added to the response model. Eisenhower *et al* (1991) described the step of encoding, which involves the formation of memory or the creation of records. Edwards and Cantor (1991) elaborated on the encoding and retrieval steps as well as discussed respondent selection.

Sudman, *et al* (2000) proposed a hybrid model based on Edwards and Cantor's research as well as

Tourangeau's traditional cognitive response process. Their eight-step cognitive response model for establishment surveys included the following:

1. Encoding of information in company records
2. Selection and identification of the respondent(s)
3. Assessment of priorities
4. Comprehension of the data request
5. Retrieval of relevant information from existing company records
6. Judgment of the adequacy of the response
7. Communication of the response
8. Release of the data.

These steps account for circumstances unique to the establishment survey setting, specifically relating to:

- the use of records, as opposed to memory recall, when completing a survey request,
- distributed knowledge, which affects who is chosen as a respondent, and the number of people required in order to respond appropriately to a survey,
- competing priorities, both for the organization as a whole and for the individual respondent(s) completing the survey, and
- authority for data release, in which only some members of the organization are authorized to release the data that has been reported on the survey.

Electronic reporting options have grown since the Census Bureau first introduced electronic reporting for economic surveys in the late 1980s. In the beginning, only a select number of companies were invited to report electronically. The software and the internal infrastructure to support the software was not sophisticated enough to support mass distribution. Currently, most surveys that have an electronic reporting option are open to all respondents. Over time the user interface of the software has improved, along with the addition of new functionality to assist users in their response task (Sedivi, 2000). In the late 1990s the Census Bureau introduced its first survey administered via the Web.

Surveys administered electronically have some advantages over their paper counterparts. These advantages are due to certain functionalities that can be built into the software to assist respondents in reporting more accurate, and sometimes more timely, data (Sedivi, 2000). At the Census Bureau, every electronic survey administered to businesses, whether via software or the Web, includes edits. These edits check the respondents' data for missing or inaccurate values. Respondents have the opportunity to update

and correct their information or provide an explanation for why a questionable value is correct. [1] Edit messages help a respondent provide more accurate information and lower the respondent's burden associated with follow-up contact that often occurs when questionable data are detected. Another advantage of some electronic surveys is the option to compile data in a spreadsheet and then "import" that data from the spreadsheet into the software, thus sparing the respondent from typing in massive amounts of data. This importing feature is typically used by large and medium sized companies that have to provide detailed data for each of their locations.

3. Methodology

The research undertaken by the U.S. Census Bureau to develop user requirements for Surveyor provided the vehicle to explore the effect of electronic instruments on the establishment response process model. Surveyor is the software used to collect data from business respondents on an annual or quinquennial (5-year) basis. Here we describe the methods we used to gather those user requirements.

3.1 Panels of Respondents

Detailed user requirements were necessary in order to make improvements to the existing electronic data collection software, which had been originally developed for the 2002 Economic Census. We gathered these requirements through the use of two panels of respondents – a longitudinal panel and a rotating panel.

Respondents in the longitudinal panel were visited multiple times, on average about once per calendar quarter. Each visit addressed a different area of the software, including edit messages and the help section. We were able to build upon information provided at earlier meetings, and engage in a more in-depth discussion of the issues at hand.

The rotating panel members were visited only once during the research process. While we thought it was important to maintain a connection and build rapport with respondents through the longitudinal panel, we did not want to lose the opportunity to get a "fresh look" at the software from other respondents who did

---

[1] Missing values and data with inconsistent values receive edit messages that prompt the respondent to make changes to possible erroneous data. The edit messages contain the location of the possible problem and a description of what actions are necessary to resolve the problem.

not work with us, or the software, on a regular basis. Respondents in the rotating panel went through a series of questions and tasks similar to those in the longitudinal panel, and they also provided information related to their usual response process.

3.2 Task Analysis

Early on in the process, we asked our respondents to describe how they went about reporting data to the Census Bureau for the 2002 Economic Census. We asked many questions, concerning a great number of topics, including the following:

- The types of people involved, what departments they represented, and what type of information they provided,
- Where data was kept,
- Who had access to the data,
- If the respondents' technical support staff got involved, and
- How they moved data from their own records and systems into the software.

We were able to obtain more information about the response process from the longitudinal panel members for a couple of reasons. First, we spent a significant amount of time at our initial meetings covering the topic. Second, on subsequent visits, we were able to further elaborate and probe the issue. Because we had only one meeting with each of the rotating panel members, the information from them was not as detailed.

3.3 Respondent Exercises

In addition to the analysis of tasks performed by respondents in their electronic reporting history, we wanted to observe how they would currently perform tasks. The best way for us to do that was to ask respondents to complete certain exercises using the software. The exercises gave respondents the opportunity to provide specific and concrete feedback about an activity, rather than speak abstractly about the actions they had taken in the past or the actions they would take using the software.

Exercises were completed in the researchers' presence, during the visit, and were designed so that respondents had specific tasks to accomplish, while providing an opportunity for the researcher to make observations of the respondent. Through these exercises, we were able to observe how respondents chose to navigate within the software and how they transferred data from their systems into the software.

At the beginning of the project's life, we talked about sending exercises to respondents prior to our meetings with them so that they could provide feedback about their experiences, highlights, and frustrations to us during the visit. However, given the breadth of industries and patterns of reporting among the respondents, this task quickly became too time-consuming. We dropped the pre-meeting exercise technique in favor of compressed exercises conducted during the visit.

There were occasions when we sent material to respondents in the longitudinal panel and asked them to read it prior to our meeting. No interaction with the software was necessary; this task solely involved reading. Asking respondents to read something in our presence, during a meeting, had proven time-consuming and rather intimidating for respondents, so they did not attend to it as well as if they had read it on their own, before our arrival.

3.4 Prototypes

In cases where we were unable to have fully functional software made for the purpose of getting respondent feedback, we developed low-fidelity prototypes. These prototypes enabled us to demonstrate how the instrument would behave and what it would look like, providing a concrete example for respondents to comment on.

In one instance, we guided respondents through a series of paper screenshots and described what would happen with each click of the mouse. In another prototype, we took respondents through a PowerPoint mock-up of what the software would do and what it would look like.

3.5 Respondent Preferences

While prototypes and exercises can address the issues of functionality and navigation, issues of formatting can be better addressed in different ways. We began by asking respondents to describe their ideal instrument and then asked questions about the specifics of what it would look like. For example, "when the software creates a spreadsheet, and the expected answer to the question is a yes or no response, how would you like to communicate that to us?" ("y" or "n," "yes" or "no," "0" or "1," etc.?). These specific questions led to discussions about what respondents wanted in a spreadsheet and how spreadsheets would impact their response process. We probed on topics such as the type of information that should be included in column titles and how to relay instructions. Later visits with members of the longitudinal panel built

upon the information gathered in earlier visits, as we constructed our spreadsheet based on what respondents had told us.

4. Results

While our research provided a good basis for the user requirements specific to the software being developed, it also provided further insight into Sudman *et al*'s hybrid response process model for establishment surveys. In this section, we return to the eight-step model outlined in Section 2 and discuss the implications of electronic instruments on selected steps within that model.

4.1 Respondent Selection and Identification

Most respondents to the economic census who are in medium and large sized companies must go to others within the company to gather data. It is uncommon for a respondent to have access to all of the data requested. During the course of our research, we discovered that there are several factors that affect the respondent's choices of methods for data retrieval from others. These choices were typically based on the internal structure of the company and the preferences of the respondent. Respondents decided what they thought would work best with their existing computer knowledge, the data itself as well as its structure, their data gathering procedures, and the internal company structure.

Retrieving data from others within the company in order to complete the survey often involves the use of spreadsheets. In the software used for the economic census, the primary respondent can either create a spreadsheet for distribution using the software or use a spreadsheet supplied by internal contacts and "map" it to the economic census software. Respondents who distribute a spreadsheet must also decide their method of distribution:

- Parse the spreadsheet by rows or columns, sending only the necessary pieces to each contact, or
- Provide the entire spreadsheet to each contact, and instruct them to complete only certain parts.

Upon receiving data from their internal contacts, the respondent must then determine how to import the data into the software. Respondents can import each spreadsheet into the software on a piecemeal basis or combine them into one master spreadsheet. If they choose a master spreadsheet route, then respondents must decide whether to cut and paste data from individual contacts into a blank master or append rows/columns to the master as spreadsheets are received from their internal contacts.

4.2 Comprehension of the Data Request

4.2.1. Moving from a piece of paper to a software program.

In addition to the instructions related to specific questions in the survey, respondents must also work with instructions specific to electronic reporting. Electronic instruments require more mode-specific instructions than paper instruments. Completing a paper questionnaire is often a straightforward task: respondents use a pen or pencil to check boxes, fill in circles, or write numbers or words into specified answer boxes. Upon completion, the respondents return the paper form in the enclosed envelope. Completing that same survey using software requires a different knowledge base that can vary depending on the user's level of computer experience.

There are several steps required for completing a survey electronically which are not required of its paper counterpart. Some of these steps include:
- Determining if a computer meets the software's minimal system requirements,
- Downloading the software (sometimes requires user IDs and passwords),
- Locating the software on the computer's hard drive after it has been downloaded,
- Viewing and manipulating multiple survey forms,
- Importing and exporting (if respondents choose to use that functionality),
- Opening an individual survey form,
- Navigating through the form,
- Working with and understanding the edits and their messages,
- Locating and navigating the help section to find necessary information, and
- Submitting the data electronically.

4.2.2. The savvy user versus the non-savvy user

The ability of the respondent to comprehend all of these steps – some necessary, some optional – depends on the user interface of the software, the clarity of supporting help material, and the computer ability of the respondent. Survey institutions can improve the design of the user interface and the clarity of the help material through research. The computer ability of the respondent is a varying factor.

Developing a user interface and help material easily understood by respondents regardless of their computer ability is a challenge. Not providing enough information confuses less savvy respondents and may deter them from responding electronically. In our research, we found that these types of respondents spent a great amount of time trying to understand instructions. In some cases, these respondents did not realize that some useful and time-saving functionalities were incorporated into the instrument.

4.3 Retrieval of Data

Respondents who choose to gather their data via spreadsheets and take advantage of the import functionality built into the software must initially decide how to retrieve that data from their internal systems. Sometimes this entails the involvement of internal Information Technology (IT) staff who maintain these systems.

The involvement of IT staff in the data retrieval process is not always necessary within companies, depending on the systems within the company (how they are set up and what data resides within them), as well as the familiarity of the respondent with those systems. One company noted that they had pulled data using a more manual process that they were hoping to eliminate by involving IT staff who could script programs to pull the data together more quickly.

Some companies require a detailed review of all software by IT staff before permission is granted to download it. The involvement of IT staff – whether for data retrieval or software review – adds another complication to the response task.

4.4 Judging the Adequacy of the Response

4.4.1. Edits

All of the electronic instruments designed for establishment surveys at the Census Bureau have edits. These edits check for data inconsistencies and prompt the respondent to fix or comment on anything that falls outside the edit parameters. This direct interaction between respondents and an edit is unique to self-administered electronic instruments. In interviewer-administered surveys, respondents only deal with edit messages indirectly, through the interviewer. In self-administered non-electronic surveys, respondents do not deal with edit messages at all.

When respondents work with edit messages, they make judgments about the validity of the response that they reported. They determine whether that value is indeed incorrect or if that value is correct for their company (and out of range of the software's edit parameters). When the disputed value is correct, respondents are encouraged to give an explanation about that value. Although this process adds burden initially, the explanation helps analysts reviewing the data to understand why the value is correct and helps to prevent follow-up contact with the respondent that sometimes occurs when values fail an edit after data are received.

4.4.2. Reviewing data for accuracy

Respondents typically review their data prior to submission, checking to make sure it is accurate. Respondents might spot-check values or compare totals from the requested data to totals available in the financial records of the company. Not all respondents for an economic census can review data for accuracy. Because of the many different layers that some respondents must go through to gather the data for an economic census, the person responsible for submission of the data may not be familiar enough with the data provided in order to review it. In these cases, the coordinating reviewer must rely on their colleague's knowledge of the data provided.

4.4.3. Reviewing data prior to importing

In addition to reviewing data for accuracy, respondents who choose to use spreadsheets with the software must ensure that the data they provide meets the software's expectations for importing. For example, respondents must ensure that the spreadsheet contains the appropriate and expected values in each column. In addition, a respondent must make sure that the field does not exceed the specified length. The consequences of the spreadsheet not meeting the software's expected format include data truncation, triggered edit messages, or import failure.

4.5 Reporting the Response

4.5.1. Data formats

Organizations that report data electronically must ensure that their data fit the format that the software is expecting. For instance, the components that compose a date – day, month, year – have to be in the appropriate order and (not) contain hyphens or slashes. Either the data fits the software's expected format, or the respondent must make corrections when edit messages appear.

4.5.2. Saving records

Both paper and electronic respondents are encouraged to save a copy of their forms. Respondents to business surveys, paper or electronic, typically maintain a copy of their submitted responses for the record. When dealing with electronic forms, respondents must decide how they should keep these copies. Some respondents prefer to print off hard copies from the electronic forms to keep in their files. Some respondents prefer to keep electronic copies of their responses. Companies that use the spreadsheet functionality tend to maintain a copy of that spreadsheet in their records. In addition to copies of the forms, paper or electronic, respondents also keep copies of any supporting material that was used in gathering the data. Respondents want a record of how the data was gathered so that they can repeat that process in the future or have it documented for the next person assigned to the task.

4.5.3. Submitting data electronically

Submitting data to the Census Bureau electronically through the software is a different process than submitting paper forms through the mail. While paper forms merely require an envelope, the electronic reporter must follow a series of instructions and prompts before successfully submitting their data. Respondents have noted in the past that there is more anxiety associated with submitting data electronically than through the mail, as there are concerns about data getting lost in transmission. In response to these concerns, the Census Bureau created a survey status page on the Internet during the 2002 Economic Census that allowed respondents to log in and verify that the Census Bureau had received their data.

5. Discussion

In a paper-based establishment survey environment, respondents interact with the question, their records, and the paper instrument. Paper instruments, when designed appropriately, require little effort on the part of the respondent, since individuals in the US read from top to bottom, left to right, then turn the page to continue. However, electronic survey instruments, despite their foundation in known software environments (e.g., Windows, Internet), are new enough that few design conventions exist and are followed. As a result, respondents interact with an instrument they must seek to understand. They are therefore interacting with the question, their records, and an electronic instrument.

Perhaps as electronic survey instruments become more widely used and available, they will begin to take on a set of conventions all their own. When that day comes, and when electronic instrument designers apply those conventions uniformly and consistently, survey data collection software will begin to be like paper is today – a known instrument that people can move through with ease and comfort.

In a paper survey, respondents interact with the question and with the design and layout of the questions on a page. In an electronic survey, respondents interact not only with the questions and their design, but also with the navigation capabilities and the built-in functionality. As the number of interactions increases, so does the risk for error. To the extent that one of those interactions is with an unfamiliar environment, as is the case with electronic survey instruments, especially stand-alone software, we risk causing confusion and frustration among data providers. How respondents handle edit messages, as well as navigation through the software, can add cognitive burden to the response process, making the process more difficult. As data collectors who rely on respondents for information, it is our responsibility to ensure that we do not place undue burden on this process.

Developing user interfaces, instructions, and functionalities that appeal to both savvy and non-savvy computer users is a topic for further study, as we continue to learn more about our users and how they interact with electronic surveys. We have shown the impact of having an interface designed with the savvy user in mind: less savvy users struggle or miss important functionalities and features, they spend excessive amounts of time digging through help instructions to operate the software, and they spend more time on the telephone with our technical support staff. Sometimes users abandon the task entirely, a phenomenon of which we have no true measure of. Topics for future research include:

- Would designing a user interface for less savvy users have any negative impact on savvy users?
- Is there any way to accurately gauge the number of users who abandon the survey after downloading it because of issues related to usability?
- Is there any way to determine, possibly through an event log, which parts of the software respondents have problems with?

Some electronic surveys at the Census Bureau are very similar in layout and design to their paper counterparts. The subject matter specialists in charge of those surveys often make electronic form design decisions. The debate continues, even within the Census Bureau,

about how closely a paper form should resemble its electronic counterpart. We have found evidence that for longer and more detailed business surveys, respondents print out the paper form to use as either a 'scratch sheet' or a guide for how to gather the data on the survey. What remains unknown are the potential mode effects that are introduced when the electronic form differs from the paper form. Further study should be given to the effects of seeing the same survey in paper and electronic form. In addition, there has been little research examining possible mode effects associated with different methods of electronic reporting. Perhaps there is a significant mode effect between electronic reporting via software versus via the Internet.

Allowing respondents to use spreadsheets in conjunction with the electronic data collection software at the Census Bureau has raised other error concerns arising from the mode. When collecting data via spreadsheet, respondents often put question wording and instructions to the side and/or top. When spreadsheets are passed from person to person within a company, the question wording and instructions do not always get transferred. Sometimes respondents are re-stating the questions or truncating the question wording, which raises concerns about whether the data collected is the same as the data requested.

Finally, we posit that electronic reporting affects the traditional cognitive response process model for establishment surveys in the following ways:

1. Respondents must comprehend more than the verbal and the visual language. They must comprehend the instrument itself – its features and functionalities, and its capabilities and limitations.
2. Electronic reporting affects how data are gathered from other individuals within the organization, especially with regard to how data are transferred from internal systems.
3. Respondents work with edit messages directly, a feature unique to self-administered electronic reporting. Furthermore, respondents conduct an additional review, ensuring that the data meets the software's requirements.
4. Saving the data for internal record-keeping purposes and transmitting the data should be added to the response model, though they are not unique to electronic reporting.

6. REFERENCES

Edwards, W.S. and Cantor, D. (1991). "Toward a Response Model in Establishment Surveys." In *Measurement Errors in Surveys*, P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (eds). New York: Wiley.

Eisenhower, D., Mathiowetz, N.A., and Morganstein, D. (1991). "Recall Error: Sources and Bias Reduction Techniques." *In Measurement Errors in Surveys*, P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (eds). New York: Wiley.

Sedivi, B., Nichols, E., and Kanarek, H. (2000). "Web-Based Collection of Economic Data at the U.S. Census Bureau." Paper prepared for presentation at the Second International Conference on Establishment Surveys, Buffalo, NY.

Sudman, S., Willimack, D.K., Nichols, E., and Mesenbourg, T.L. (2000). "Exploratory Research at the U.S. Census Bureau on the Survey Response Process in Large Companies." Paper prepared for presentation at the Second International Conference on Establishment Surveys, Buffalo, NY.

Tourangeau, R. (1984). "Cognitive Sciences and Survey Methods." In *Cognitive Aspects of Survey Methodology*, T.B. Jabine, M.L. Straf, J.M. Tanur, and R. Tourangeau (eds). Washington, DC: National Academy Press.