The Paradox of Respondent Burden in Testing Electronic Instruments for Establishment Surveys

Diane K. Willimack[*]
U.S. Census Bureau

1. Introduction

The pretesting of establishment surveys faces several challenges related to the general nature of data collection from businesses or organizations (Willimack et al., 2004). Data requested from businesses are often technical, requiring precise definitions and specialized reporting instructions. Data are also expected to reside in business records. However, data necessary to meet a single survey may be distributed across multiple records systems, with different business personnel having specific knowledge of and/or access to the appropriate records. For some items, data residing in business records may differ, by definition, from requested data, and business respondents must identify and carry out reasonable estimation processes. The many activities involved in responding to a business survey – identifying data sources, delegating and coordinating data retrieval from multiple sources, gathering the data, preparing estimates, checking the adequacy of the data to meet the data request, and entering the data onto the survey instrument – constitute a labor-intensive response process. In addition, larger businesses tend to be selected for many surveys, because their participation is needed to ensure accurate economic statistics. Thus, this labor-intensive survey response process is multiplied several times over, resulting in additional respondent burden for large businesses in particular.

Certainly, one of the over-arching goals of providing electronic data collection instruments for business surveys is to reduce respondent burden. To accomplish this goal, software for electronic survey instruments should possess a user-centered design, which is one of the principles guiding software development. In a user-centered design, "a product's goals, objectives, context, and environment are all derived from the user's viewpoint, as well as all aspects of the tasks that the product supports." (Rubin, 1994).

According to usability engineers, creating a user-centered design is best achieved by involving end users in the software development process. Research activities designed to understand users' requirements include task analysis and usability testing. Task analyses are done early in the development cycle to identify the activities users need to perform using the software, and how they go about performing them (Usabilitynet.org, 2005). Usability testing is a research process by which typical users personally evaluate the ease of use of the software. To ensure that users' needs are indeed met, user testing should be conducted iteratively as the software is being developed. (Rubin, 1994).

So, consider the implications for software development for electronic data collection instruments for business surveys. The intent of the instrument is to facilitate data collection and ease

---

[*] This report is released to inform interested parties of research and to encourage discussion. The views expressed on methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau. The author notes that the specific studies described in this paper are the product of research conducted by members of the U.S. Census Bureau's Establishment Survey Methods Staff and their collaborators. She further acknowledges research leader Amy E. Anderson for permitting "work-in-progress" to be discussed during the QUEST meetings.

respondent burden. But, in order to ease respondent burden overall in the final, fully functional instrument, we should involve users – business survey respondents – in its development. However, involving business survey respondents in an iterative manner in electronic survey software development clearly adds burden.

The paradox of adding to respondent burden via pretesting in order to reduce respondent burden during production data collection is clearly illustrated in re-engineering software for collecting the USA's 2007 Economic Census.

The economic census is conducted by the U.S. Census Bureau every five years, for years ending in '2' and '7'. Data are used to benchmark the status of the U.S. economy at detailed industry and geographic levels, as well as to provide data for estimating U.S. Gross Domestic Product. To support these uses, the economic census collects detailed data on employment, payroll, and sources of revenue for each establishment, or physical location, of a business. For 2002, the economic census consisted of more than 600 questionnaire versions, tailored by industry. Response to the economic census can be very burdensome, particularly for large multi-unit companies, where detailed data must be gathered from multiple sources within the company and recorded on individual survey forms. Companies operating in multiple industries must also ensure that figures are obtained and entered onto the correct form reflecting the establishment's industrial classification.

In addition, 2002 was the first time that electronic data collection was offered to all establishments in the economic census. Experiences from electronic collection of the 2002 Economic Census suggested the need for a number of improvements to the software and identified functionality that did not work as intended. Thus, research was requested to identify business survey respondent-user requirements for re-engineering the data collection software for the 2007 Economic Census.

We faced the issue: How can we involve respondents in development and testing of data collection software – according to the principles of a user-centered design – for a labor-intensive, but infrequent, response task, like the economic census, in a manner that does not overly burden business respondents?

In addition to businesses' response processes being labor-intensive, the software itself – called Surveyor – had been designed with a fair amount of complexity intended to accommodate the intricacies of the economic census response task. Previous research experience demonstrated the limitations of one-time contacts with businesses for the purpose of conducting usability testing of Surveyor, due to its complexity. It was difficult to obtain adequate information in a single visit for the purpose of developing detailed software requirements useful to the programmers.

Our research experience had also shown that a single onsite visit with a business for pretesting purposes needed to be limited to 60-90 minutes. Although this amount of time was insufficient for obtaining adequate detail in a single interview to support research needs for Surveyor redesign, we wondered if business respondents would be willing to grant us multiple interviews of 60-90 minutes each, over time. In other words, would business respondents be willing to participate in iterative testing over the course of several months – that is, participate in a panel – although such a commitment would increase respondent burden? What, if anything, would we gain by using a respondent panel for testing survey software as it is being developed?

To demonstrate the evolution of this research strategy, I will describe two earlier studies conducted at the U.S. Census Bureau in which electronic reporting was targeted, along with research supporting software re-engineering for the 2007 Economic Census. For each study, I will describe its purpose, the research strategy, the types of findings that were elicited by this approach, and the limitations of the methodology. Note that the focus on findings will be on their nature and not their specific substance.

2. Study #1: Hak et al., 2003

Purpose: To identify 1) factors that influence the uptake of electronic reporting, and 2) opportunities for improving current electronic reporting instruments.

Research Strategy: Two Census Bureau electronic survey instruments were studied. One was a Web data collection instrument for a short, straightforward company-level monthly economic indicator survey, requesting at most seven items from respondents in manufacturing. For most survey participants, many, if not all, of the items could be directly retrieved from standard records. The second instrument studied was the Surveyor electronic data collection software being used to collect the 2002 Economic Census.

Researchers visited participating companies to observe respondents' use of the electronic instrument or to conduct retrospective debriefings about their experiences interacting with the software. In addition, for the economic census, a few companies were visited more than once, ostensibly to aid the respondent with reporting and submitting data in Surveyor. The subsequent visits enabled the researcher to observe respondent interaction with the Surveyor software during successive response steps, thereby obtaining additional information about the complex response process for the economic census. Alternatively, some companies were debriefed over the telephone about successive activities and their experiences with the software.

Types of Findings: Two types of results were obtained based on this research methodology. First, we learned that business respondents were receptive to multiple visits if they were seen as providing support. Second, among the substantive findings, we discovered aspects of respondent burden related to electronic reporting. Some were seen to reduce burden, others were considered by respondents to increase it. For the short monthly survey, many respondents found the benefits of electronic reporting to be marginal at best, often related to personal preferences. For the economic census, respondents articulated ways that electronic reporting added to burden:

- Communicating the response was considered by one respondent as requiring that "my data" be put into "your [electronic reporting] format."
- Surveyor's "mapping" function, whereby respondents could build a customized spreadsheet consisting only of economic census data items pertinent to their businesses, was seen as confusing and burdensome.
- Respondents were dismayed by Surveyor's inability to "copy and distribute" particular forms to other company units, a common practice among companies with a decentralized reporting process.

Limitations: The monthly survey was too simple to facilitate much depth in respondents' articulation of user requirements, saying simply, "you can't improve on something this simple." The economic census, on the other hand, was too complex. One or two company visits provided isolated glimpses into numerous, inter-linked, labor-intensive tasks, which could have formed the

basis for a task analysis around which to formulate software user requirements. In addition, scheduled appointments with respondents in order to observe their interaction with Surveyor in essence "staged" the observations, taking them out of context or making them somewhat artificial.

Nevertheless, the research findings were indeed useful in identifying some software features with which users were dissatisfied, along with initial indications about how respondents would like to use software in their response processes. Although the findings fell short of providing sufficient depth from which to specify programming requirements, they did form a foundation of user expectations upon which to build. In addition, experience from this study suggested that business respondents may be receptive to follow-up visits from Census Bureau staff within a short period of time.

3. Study #2: 2003 Surveyor Usability Testing

Purpose: To clarify the nature of the problems with the Surveyor software, which had been identified in the previous study.

Research Strategy: Debriefings were conducted with respondents following their use of the Surveyor software for an annual survey collecting a few items for each establishment, which could number in the thousands for large multi-unit companies. The usability of Surveyor was investigated during a single site visit, where a one-on-one personal interview was conducted with the respondent who actually used the software.

Types of Findings: This research uncovered additional information about the nature of some of the problems with the software that had been identified in earlier studies. To some degree, details began to emerge. Examples include:

- Technological terminology, such as "exporting" and "importing," was unfamiliar to many business survey respondents, who, for the survey studied, were likely to be accountants or specialists in human resources.
- Surveyor's "mapping" function, by which the respondent could build a customized spreadsheet, was not intuitive. Instead, the theme expressed by most respondents was "just give me a spreadsheet with columns for all questionnaire items."
- Not only did respondents want the ability to "copy and distribute" particular forms, they also wanted control over determining which forms were distributed to whom within their companies.

Limitations: Again, useful results were obtained. However, the details tended to be arbitrary and contextual, failing to provide a cohesive picture of how the software should function and behave overall. Information remained insufficient for well-specified programming requirements.

4. Study #3: Anderson, et al., 2005

Purpose: To gather detailed respondent-user requirements to re-engineer the Surveyor software for collecting the 2007 Economic Census.

Research Strategy: The research design utilized both a longitudinal panel of companies and a rotating panel. For the longitudinal panel, 20 companies were recruited and asked to permit

onsite research interviews up to four times, once per quarter, for a year. Thus, the research topics were divided among four rounds of testing. Each round consisted of interviews with members of the longitudinal panel, augmented by companies recruited for a single visit, making up the "rotating panel."

Multiple research methods were used to build upon one another over the course of the study. "Initiation" visits with longitudinal panelists included debriefings about respondents' past experiences with Surveyor and a task analysis of steps respondents undertake when completing the economic census. In addition, a paper prototype of a proposed new software feature was introduced and discussed.

Subsequent visits utilized usability testing of current software features. Respondents were also presented with prototypes of new or modified features, often having minimal functionality. For some research topics, respondents were asked to perform specific tasks or exercises indicative of activities they had described previously during task analyses, while researchers observed their behavior and probed their reactions.

Types of Findings: An in-depth task analysis became a by-product of ongoing contact with longitudinal panel member companies, as details about the response process were uncovered during initiation visits and built upon during topical interviews, probing, and usability tasks conducted in subsequent visits. This provided a cohesive integrated picture of respondents' needs and expectations for software functions and features that would aid their work.

Respondents' confusion over technological terminology was clarified, when it was discovered that Surveyor functions "exporting" and "mapping" amounted to "creating a spreadsheet." This language was meaningful to business respondents who were expert in the use of spreadsheets, if not technologically savvy.

In fact, the most profound results, in terms of the task analysis and the identification of software requirements, centered around business respondents' integral and efficient use of spreadsheets in the survey reporting process. Details about how spreadsheets are used and manipulated, and by whom, provide a baseline for building software requirements that meet respondent-user needs.

The continuity offered by the longitudinal panel not only contributed to the depth of the research findings. It also enabled the research team to build on past feedback, ensure a "meeting of the minds" on desired features and functionality, and re-test the result to determine its efficacy in meeting user expectations.

Limitations: While the panel approach enabled Census Bureau researchers to obtain the depth of information necessary to provide meaningful and suitably specific software requirements to Surveyor programmers, it is not without its limitations. Repeated visits with the same respondents on the same topic is akin to dependent interviewing. Results from one meeting to the next cannot be viewed as independent. Thus this approach may not be suitable for long term software testing, where the software must be usable to new users as well as those who are experienced.

The rotating panel was intended to compensate for this deficiency. While this was helpful to some degree, it also required separate protocols to provide context for rotating panel members. In addition, the panel approach with business survey respondents may be inappropriate for short-

term projects, as company reporters may not tolerate frequent visits from researchers.

A major limitation for this project was the inability to prototype alternative features and functionality within the Surveyor software itself. Using a prototype with limited functionality also potentially limits the applicability of the findings, if the desired functions or features behave or appear substantially different when programmed in Surveyor.

5. Summary

I began by posing the problem of designing a data collection software testing plan for the 2007 Economic Census, which requests detailed establishment-level information, resulting in a difficult, labor-intensive response process for businesses. The goal of creating user-centered software, which relies on user (respondent) involvement during development, seemed at odds with our experience of limiting pretest interview length with business respondents in order to control respondent burden. Nevertheless, we had some evidence that business respondents may tolerate multiple contacts over an extended period of time, if perceived to be to the business' advantage. Thus we wondered whether we could successfully recruit and maintain a "panel" of business respondents to participate in iterative software testing.

We found that businesses were willing to participate in a research panel. All twenty longitudinal panel members participated in more than one onsite company interview, and seventeen of them participated in three or four rounds of testing. Business respondents recognized the benefits to them of contributing to the improvement of software they must use if they want to take advantage of electronic reporting opportunities. They took "ownership" of their role in the development process. In addition, personal attention from Census Bureau researchers bred good-will as relationships with respondents were developed and improved.

We accrued a number of benefits from using a respondent panel approach to achieve the research goals of identifying user requirements to guide re-engineering the Surveyor software. For complex surveys, previous research based on one-time interviews with business respondents failed to obtain adequate depth and detail, regardless of whether the method used was observation, retrospective debriefings, or post-collection respondent debriefings. Because of the complexity of both the response process and the software, repeated contact was needed for researchers to fully grasp – and grapple with – respondent-user needs. Rather than propose hypothetical situations for respondent reactions, prototypes and examples enabled researchers to observe a semblance of reality of response tasks and their consequences. The continuity of the panel approach supported corroboration and validity checks as software was refined based on previous findings and re-tested in subsequent interviews to ensure a "meeting of the minds" between the researcher and the respondent.

The software development and testing process cannot end here, as the achievement of a user-centered design relies on functionality that is so usable as to be transparent to the user. Thus future testing of a re-engineered Surveyor will involve business respondents new to the testing protocol, where, hopefully, the benefits of the panel approach to identify and specify user requirements will be seen – or perhaps not seen!

References

Anderson A.E., Morrison R.L., and Brady, C.F. 2005, “Applying Knowledge of Business Survey Response Processes to the Design of Data Collection Software at the U.S. Census Bureau,” paper presented at the Federal Committee on Statistical Methodology Research Conference, Arlington, VA, November, 2005.

Hak, T., Anderson, A.E., and Willimack, D.K. 2003, “Determinants of Web Reporting: A Qualitative Study of Mode Selection,” Proceedings of the Federal Committee on Statistical Methodology Research Conference, Washington, D.C.: Office of Management and Budget.

Rubin, J. 1994, Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests, New York: Wiley.

Willimack, D.K., Lyberg, L., Martin, J., Japec, L., Whitridge, P. 2004, "Evolution and Adaptation of Questionnaire Development, Evaluation, and Testing Methods for Establishment Surveys," in S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer (eds.), Methods for Testing and Evaluating Survey Questionnaires, New York: Wiley.

Usabilitynet.org. 2005, “Task Analysis,” http://www.usabilitynet.org/tools/taskanalysis.htm. Accessed 29 June 2005.