

APPLICATION OF COGNITIVE METHODS TO AN ESTABLISHMENT SURVEY: A DEMONSTRATION USING THE CURRENT EMPLOYMENT STATISTICS SURVEY¹

Eileen O'Brien, U.S. Census Bureau, Sylvia Fisher, Karen Goldenberg, Richard Rosen, Bureau of Labor Statistics
Eileen M. O'Brien, U.S. Census Bureau, Room 3133-4, Washington, D.C. 20233-9150

KEYWORDS: Cognitive interviewing; Expert Appraisal; Establishment survey; Measurement error

1. Introduction

Cognitive survey research methods have been used successfully for more than 20 years to improve social and demographic survey questions, and more recently to improve establishment surveys. This paper describes the use of two different cognitive research techniques to evaluate the questionnaire for the Current Employment Statistics (CES) survey, a monthly establishment survey conducted by the Bureau of Labor Statistics (BLS). We look here at the relative contributions of the two methods for identifying sources of measurement error.

Section 2 provides a brief background on the CES survey and the two cognitive methods used in this research. Section 3 describes how we applied these methods. In Section 4, we report key results, while Section 5 synthesizes our findings and offers a discussion of the relative strengths of each approach.

2. Background

Cognitive research methods help to identify potential sources of measurement error in specific questions, series of questions, or questionnaire layouts. These methods, which are qualitative in nature, identify respondent difficulties and why they may occur by relating them to the four-stage cognitive response process: comprehension, retrieval, judgment, and communication (Tourangeau, 1984). Understanding the nature of difficulties respondents have with the questionnaire helps researchers to identify areas where improvements can be made.

The methods used here rely upon the response models suggested by Tourangeau (1984), Edwards and Cantor (1991), and Sudman et al. (2000). Edwards and Cantor and Sudman et al. expand on the basic response model and incorporate explicit response steps for establishment respondents, including record formation, selection of the respondent, assessment of priority for the data request, and review of data prior to releasing it.

The cognitive methodologies used in this paper draw from all of these response models. They are:

(1) an expert appraisal of the CES questionnaire, and
(2) cognitive interviews with establishment survey respondents.

2.1 Expert Review/Appraisal

In general, *expert review* is a process by which questionnaire design experts evaluate questionnaires for potential wording, navigation, or other problems that might reduce response quality. The *expert appraisal* is a formal tool that requires reviewers to systematically consider each step of the response process. It consists of a set of codes describing question features that may contribute to response error. Trained coders conduct a standardized item-by-item review of a questionnaire and assign codes to items they believe induce error. Researchers then tabulate the frequency of codes and may use that information to guide evaluation, to set cognitive testing priorities, or to suggest revisions to the survey instrument.

By providing a simple count of potential problems where they might occur, the expert appraisal augments the rich detail of a qualitative methodology with some benefits of a quantitative approach. Analysts may note why particular codes were used, making the transition to evaluation, further testing, and revisions more direct. Ultimately, the appraisal provides a consistent basis for identifying problems and making recommendations for questionnaire testing and redesign (Lessler and Forsyth, 1996).

The expert appraisal tool has been adapted to establishment survey questionnaires (Forsyth et al., 1999; O'Brien, 2000a). Drawing on the work of Edwards and Cantor (1991) and Tomaskovic-Devey et al. (1994), the establishment version includes codes that address the complexity of the organizational response process. For example, establishment respondents tend to draw more on records than on memory to construct answers. They may have to obtain complex information from multiple individuals or sources and integrate that information. Respondents' cognitive processes and work roles, aspects of the organization and its structure, and the external environment in which the firm operates also influence how a survey request is perceived and completed (O'Brien, 2000b; Tomaskovic-Devey et

¹ 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.

The opinions expressed in this paper are those of the authors and do not represent official positions or policies of the Bureau of Labor Statistics.

al., 1994). The establishment appraisal protocol includes codes to identify design problems related to these processes.

2.2 Cognitive Interviews

Cognitive interviews are semi-structured personal interviews conducted one-on-one with members of the survey target population. These in-depth interviews are designed to elicit insights about the "thinking process" that respondents use when answering survey questions (Forsyth et al., 1991). Unlike structured survey interviews, cognitive interviews allow respondents the freedom to elaborate on topics they find important, and give the interviewer comparable freedom to extensively explore these topics. Since cognitive research on establishment surveys is often conducted at a respondent's work site, the interviewer can also observe the real-world setting in which questionnaires are completed, including the availability and use of records.

2.3 The Current Employment Statistics Survey

In this research, we applied the cognitive methods described above to an evaluation of the CES survey questionnaire. CES is conducted monthly by the BLS using a sample of approximately 350,000 business establishments, and is the source of data on U.S. payroll employment, hours, and earnings, by industry. The survey program collects five data elements each month from participating establishments: *All Employees*; *Women Employees*; *Production Workers* (in goods-producing industries) or *Nonsupervisory Workers* (in service-producing industries); *Production Worker Payroll* or *Nonsupervisory Worker Payroll*; and *Production Worker Hours* or *Nonsupervisory Worker Hours*. In addition, CES obtains *Overtime Hours* for production workers in manufacturing industries. The reference period for all data elements is *the pay period that includes the 12th of the month*.

CES is a multi-modal survey. It collects data from some respondents by mail or fax and some using Computer Assisted Telephone Interviewing (CATI). The survey obtains most of its monthly data by Touchtone Data Entry (TDE), where respondents self-report by calling a toll-free number and entering monthly data from their telephones. New respondents are enrolled through telephone (CATI) interviews. Most participants respond via CATI for a few months and then convert to TDE. CES is also beginning to use the internet for data collection.

Regardless of mode, the questionnaire or data collection form provides a vehicle for compiling information. Each January, respondents receive new forms on which to record their data. Forms are pre-printed with the name and address of the business and

the respondent, the location for which data are being requested, and the previous month's employment, payroll, and hours figures at that location.

3. Methodology

3.1 Cognitive Methods Applied to the CES

We began our research with parallel activities. One set of researchers conducted an Expert Appraisal of the CES questionnaire, while another developed a draft interview protocol for cognitive interviews (Tucker et al., 2000). The appraisers were a BLS researcher who was new to this methodology, and a researcher from the Census Bureau who had contributed to adapting the methodology to establishment surveys. The specific CES form appraised was the BLS-790H, the version used to survey firms in service-producing industries. Using the Forsyth et al. (1999) scheme, the researchers conducted a systematic review of 38 elements, including both questions and instructions. The purpose of using the appraisal method was to alert researchers to likely sources of response problems (questions, instructions, or graphics) and possible reasons for them (structure, layout, terminology, or reference period issues). These results could then be used to refine the draft cognitive interview protocol.

As noted, the cognitive interview protocol was developed separately. From the cognitive interviews, we hoped to learn where and why response problems might be occurring with the current CES form, where changes in instrument design might improve data quality, and where the task could be made easier, thereby reducing response burden. In addition, we sought information about respondents':

- General impressions of the CES questionnaire
- Understanding and use of important terms
- Approach, procedures, and records used to complete the form
- Perceptions and opinions of CES instructions, questions, form design, layout and format; and
- Preferences for future response modes.

To achieve these goals, the protocol included both general questions and structured retrospective probes, and also allowed for flexibility in probing to pursue topics and opinions that seemed salient to respondents. The research team modified the draft protocol to include the results from the expert appraisal, as well as feedback from subject experts at BLS and the design expertise of this research group.

3.2 Participating Establishments

We conducted five cognitive interviews with establishment respondents in service sector industries in the Washington, DC metropolitan area. Respon-

dents were selected based on their Standard Industrial Classification code and their availability during the interview period. They were recruited by telephone and asked to complete a faxed copy of the CES form prior to their appointments. We asked that they report two months of data. One BLS researcher conducted all five interviews.

Responding establishments ranged in size from 23 to 125 employees. Some had large cadres of part time as well as full time workers, and one firm's employees earned bonuses or commissions as well as salaries.

4. Results

This section summarizes our evaluation of the CES questionnaire based on the expert appraisal and cognitive interview results. These qualitative results serve only to illustrate the variety and nature of response problems identified among the firms we interviewed.

4.1 Contributions of the Appraisal Method

The CES questionnaire relies on respondents finding and using complex sets of instructions to orient them toward specific tasks and the intended meaning of terms. These instructions appear in several places on the form. The major concern identified through the appraisal was the absence of a clear navigational path through the form, resulting in a lack of formal control over what respondents read and when they read it. Other concerns relate to the content of instructions and inadequate definitions of important terms.

One problem identified in the current design relates to the placement of important information. For example, terms are sometimes used in questions before they are defined. Also, a single concept can have multiple and seemingly different definitions, and these definitions might appear in different places. In addition, the form uses some vague or technical terms without any specification, e.g. "establishment" is not defined. Therefore, we would expect respondents to vary in the way they interpret and use these terms, and that these variations would affect data quality.

The appraisal is constructed to identify a broad range of problems that design choices might induce. The bulk of the problems identified tend to be problems of comprehension. Often, however, it is not clear where and how such comprehension problems will affect response quality. Here, cognitive interviews offered evidence of real design limitations.

4.2 Contributions of the Cognitive Interviews

A number of general themes emerged during the cognitive interviews that either did or could have

caused errors in completing the CES form. Overall, we found that respondents had more semantic than task problems with the questionnaire, but these problems did not deter them from completing the forms.

One of the strengths of the expert appraisal is to identify problems associated with the complex information retrieval process in organizations. From our appraisal, we hypothesized that respondents in complex firms might have difficulty accessing and using records to complete the form. However, we did not observe this aspect of the process in our interviews, perhaps because we visited only small or medium sized firms that did not have overly complex payroll structures. Also, respondents completed the form prior to the cognitive interviews and provided retrospective accounts of their retrieval processes.

In these interviews, the authority, capacity and motivation to respond (Tomaskovic-Devey et al., 1994) resided in one person, the respondent. While this could be a function of our selection process for the cognitive interviews, we saw little evidence of organizational impediments affecting individual responses or response strategies. We also did not see hybrid record keeping practices, i.e. use of paper-based and electronic records, and issues related to retrieving centrally versus locally stored records. We expected more respondents to have problems isolating data that conformed to the survey reference period. Difficulty using reference periods may help respondents decide whether to use records, memory, or rule-based estimation strategies.

We learned from cognitive interviews that response strategies varied based on the nature of the business and the data element in question. For example, respondents used a counting technique to construct their answers if the question requested a rare item, e.g. women employees in one firm, especially if accessing records required extra work. Where consulting records was not burdensome, or where exact figures were available in records, respondents used records rather than constructing answers from memory.

Structured probes helped us to identify problems with specific terms. In this case the cognitive interviews were more useful than the appraisal, because these probes pinpointed the specific nature of the problem. Several errors arose because respondents skipped the instructions. In fact, two of the respondents believed the form was so easy that they did not read any of the instructions. They committed several reporting errors as a result. For example, one respondent failed to correctly include the owner in the *all employee* count.

Some comprehension errors occurred where respondents interpreted words incorrectly or made improper judgments about terms used on the form.

These terms are not necessarily well defined in the instructions, even if the respondents had referred to them. For example, there was uncertainty about whether certain job classifications fell within the “nonsupervisory employee” category. One case involved “team leaders,” who are paid a higher rate than other employees in their job category, but who do not have hiring and firing responsibilities. Clearly, a more explicit definition of what constitutes “supervisory” duties and which categories of employees to include (e.g., regular, temporary, contract, etc.) would help respondents to decide how to count nonsupervisory employees. Also, respondents should be told explicitly on the form to consult the definition of the term “nonsupervisory employees” before providing their answers.

In addition to problems associated with the content of instructions, instruction formatting and layout were not conducive to comprehension. The appraisal suggested that problems could arise from the complex graphical structure and the nonlinear layout of important information. The cognitive interviews revealed the effect of layout on respondent behavior and subsequent error.

A self-administered questionnaire needs a navigational guide to direct respondents through the form, ensuring that they do not bypass important information (Dillman, 2000). The CES form does not have such a path through the questions and instructions. Indeed, four of the five interviewees reported that, upon receiving the form, the first thing they noticed was the large grid in the middle of the front page of the form. Several noted that they were drawn to the grid, because, as one said, it was where it appeared “the work would be.” These respondents often overlooked the information that frames the items in the response grid, and never saw the instructions and definitions printed on the back of the form (the second faxed page).² Furthermore, none of the respondents mentioned concerns about the absence of general instructions or orienting information at the top of the page, reporting that an “orientation section” was unnecessary.

Respondents often overlooked an entire column in the grid, thinking it was for office use. Its purpose, denoted in instructions on the form's back, is to collect reasons for unusually large changes in employment, payroll, or hours. This data element is intended for internal use and is not tabulated or

published.

The respondents demonstrated several problem-solving approaches to the form, particularly in the retrieval and judgment steps. All five respondents used computer-generated payroll reports as the basis for completing the grid, but these reports did not always contain data that exactly matched the elements requested. For some items, data from the reports had to be manipulated, introducing a judgment process in the form of a computation. For example, all five respondents determined the number of women employees for Column 2 by counting, based on memory. Similarly, most used a “subtraction strategy” to calculate the number of nonsupervisory employees, the nonsupervisory employee payroll, and total nonsupervisory hours, an approach which also requires judgment about who is a nonsupervisory (or supervisory) worker. Respondents noted that they could have produced new reports from an automated system, but indicated that these manipulations were easier to do by hand because of the relatively small numbers of employees in their firms, all of whom they knew. Of course, manual manipulations are also subject to computation and other errors.

In one notable case, information for nonsupervisory employee hours did not appear in records at all and could not be derived from records. The reported figure was highly inflated because the automobile service/repair personnel were paid in billed labor hours, rather than actual clock hours. It is unclear how pervasive this practice is within this industry or across other industries and how this might affect the reported data.

Overall, the respondent comments and behaviors observed during the interviews provided evidence of numerous difficulties with the form at various stages of the cognitive response process, particularly with regard to the comprehension, retrieval and judgment steps. We saw only minimal problems relative to the communication step, where respondents actually record responses on the form. Regardless, these difficulties suggest that a number of errors may result from the current CES design. Cognitive interviews were useful in identifying some of these problems.

In the cognitive interviews, we asked respondents for their general impressions of the form and the response tasks. From this, we got a sense of how the survey request is perceived by the respondent. As noted, most said that they were drawn first to the complex grid, and were undeterred by the lack of information on where to begin, what they should do next, and whether they were done. They also felt they had successfully completed the task.

Four of the five respondents initially thought the CES questionnaire appeared intimidating, due to the

² The CES questionnaire is a one page form, printed front and back. Because we faxed the form to respondents in advance of the interview, they actually saw a two-page document. Some of the behaviors observed here might not be the same if respondents had been looking at actual printed forms. (On the other hand, the CES program regularly faxes copies of the form to respondents, so the experience reported here remains relevant.)

print size, the volume of information, the appearance of the grid, etc. All of these respondents also commented later that the task turned out to be easier than expected. This does not mean, however, that they went about it correctly, nor does it mean they understood everything on the form.

5. Conclusions

What do the two cognitive methodologies tell us about the CES questionnaire? The appraisal system and the cognitive interviews are complementary tools. While they can be, and often are, used independently, we conducted the appraisal both to evaluate the questionnaire and to identify issues that should be explored in depth in the cognitive interviews. The interviews provided an assessment of whether the potential issues were real problems, and if so, how significant those problems were. In the current context, some of the potential problems found through the appraisal turned out not to be very important for the interviewed respondents, but still could affect the quality of CES data.

All data collection involves tradeoffs between cost, time, and accuracy. It appears that the CES questionnaire was designed to minimize cost and time. The form provides an efficient format for recording data, for reporting under a wide variety of collection modes, and for easy data entry. The cognitive tools suggest, however, that these operational advantages may lead to measurement error: the difference between what is actually reported to CES and BLS definitions for each of the data elements. The cognitive methodologies identified four features of the form that we believe contribute to this error: (1) the position of the instructions on the back of the page; (2) the technical terminology used, which might not be understood as intended; (3) mismatches between information stored ("encoded") in respondent records and the definitions used on the form; and (4) the overall graphical layout of the questionnaire. This list is only illustrative and does not cover all of the potential sources of error suggested by the cognitive tools. It also reflects the limitations of such a small number of interviews.

5.1 Relative Benefits of the Appraisal Method

With the expert appraisal tool, a researcher can inexpensively scrutinize some or all elements of a questionnaire. Using this tool may reveal potential sources of error. It can identify a full range of possible cognitive problems, even though interviews with respondents might uncover only a few of them.

One weakness of the expert appraisal is that problems emerge from an expert's perspective. As this research indicates, respondents are differentially equipped to overcome weaknesses in a survey design.

Thus, problems identified by experts may actually seldom or never occur among respondents, while respondents might encounter other difficulties not anticipated by the expert appraisal.

Another weakness is that the appraisal method offers a microscopic view of limitations within a particular survey. By focusing on individual questions, it does not identify errors that may span a series of questions or the questionnaire as a whole. For example, there are no codes to alert the reviewer to possible context or order effects. These are often detected only by taking a broader view of a questionnaire, relying on previous experience with similar items from other surveys, finding evidence from the literature, or from field observation.

Finally, the expert appraisal method weights all problems and concerns equally. However, various types of problems could affect data quality in dissimilar ways. Errors could fall in different directions, magnitudes, or relative importance. As a result, we base our decisions on where to focus our redesign efforts on more than the appraisal results. We rely on experience with a variety of response problems, some which is formally documented in the survey research literature, and bring in practical considerations.

The expert appraisal is often a first rather than the only step in a redesign effort. It would be beneficial to document the relative influence of the appraisal on subsequent research steps. For example, it may help to focus the scope of more expensive methods such as cognitive interviews.

5.2 Relative Benefits of the Cognitive Interview

This research shows that cognitive interviews best reveal where problems actually do occur, which helps researchers identify more precisely where improvements should be most beneficial. In addition to problems with particular terms or questions, respondent behavior often shows how they arrive at these trouble spots and how they resolve (or fail to resolve) them. A researcher can observe the response process as respondents address questions, instructions and layout and formulate their answers.

The weaknesses of the cognitive interview method relate to its limited scope. Because interview time is limited, a cognitive interview protocol typically focuses on a subset of potential error issues, e.g., terms, format, instructions, layout, etc. Some serious problems may go undetected if they are not explicitly probed, or if they do not apply to the current respondent. Limited or non-representative samples may bias findings, and may mislead us to believe that no problem exists, or conversely that a problem is ubiquitous across similar respondents within a target population when in fact it affects only a few. Furthermore, it may be inappropriate to gener-

alize findings from interviews conducted only with small firms when the survey ultimately serves firms of widely varying sizes and organizational complexity. Since large firms house more complex information infrastructures, we expect entirely different response issues to arise that affect data quality. Some surveys, including the CES, modify data collection procedures for large firms accordingly. Without explicitly inquiring across different sized firms, it is difficult to understand the organizational response process in which the respondent is embedded as well as the interaction between the respondent and the organization as it relates to response quality.

5.3 Recommendations Based on this Research

In reviewing the strengths and weaknesses of the two methods, we recommend that a test plan for an establishment survey first take advantage of the full coverage offered by the expert appraisal tool. We believe that the appraisal results are a useful way to make more efficient use of cognitive interviewing resources, since they can help to target potential problem areas. We also suggest that samples drawn for cognitive interviews be stratified by size of firm and industry type, in keeping with the research objectives. This stratification would contribute to a fuller understanding of both respondent and organizational response issues. This approach will allow us to observe where respondents rely on memory or where they need to or prefer to use records. It will show where respondents try very hard to answer, and where they expend little effort. It will also demonstrate areas where respondents spontaneously offer comments about heavy response burden, about a lack of motivation, or where they verge on leaving an item blank and the reasons why. Future research might also address the influence on the response process of the social environment within firms of different sizes. This might lead to a more accommodating design and data collection strategies for establishment survey respondents.³

We also suggest that subject matter experts be included at each developmental stage of the research process. Establishment questionnaires are intrinsically complex. Involving such experts should ensure that research driven recommendations are technically appropriate.

References

³ We note that the CES form is currently being revised. The new design addresses research findings from this and other studies.

Biemer, et al. (eds.), *Measurement Error in Surveys*, New York: John Wiley & Sons, 211-233.

Forsyth, B. H. & Lessler, J. T. (1991), "Cognitive Laboratory Methods: A Taxonomy," in P.P. Biemer, et al. (eds.), *Measurement Error in Surveys*, New York: John Wiley & Sons, 393-418.

Forsyth, B. H., Levin, K., & Fisher, S.K. (1999), "Test of an Appraisal Method for Establishment Survey Questionnaires," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 145-149.

Goldenberg, K. L. & Stewart, J. (1999), "Earnings Concepts and Data Availability for the Current Employment Statistics Survey: Findings from Cognitive Interviews," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 139-144.

Lessler, J. T. & Forsyth, B. H. (1996), "A Coding System for Appraising Questionnaires," in N. Schwarz and S. Sudman (eds.), *Answering Questions*, San Francisco: Jossey-Bass, 259-291.

O'Brien, E. M. (2000a), "A Cognitive Appraisal Methodology for Establishment Survey Questionnaires," *1999 Federal Committee on Statistical Methodology Research Conference: Complete Proceedings*, (pp. 307-316). Statistical Policy Working Paper 30, Washington, DC: Office of Management and Budget.

O'Brien, E. M. (2000b), "Respondent Role as a Factor in Establishment Survey Response," *Proceedings of the Second International Conference on Establishment Surveys, American Statistical Association*.

Sudman, S., Willimack, D. K., Nichols, E. & Mesenbourg, T. L. (2000), "Exploratory Research at the U.S. Census Bureau on the Survey Response Process in Large Companies," *Proceedings of the Second International Conference on Establishment Surveys, American Statistical Association*, 327-338.

Tomaskovic-Devey, D., Leiter, J. & Thompson, S. (1994), "Organizational Survey Nonresponse," *Administrative Science Quarterly*, 39, 439-457.

Tourangeau, R. (1984), "Cognitive Sciences and Survey Methods," in T. Jabine, et al. (eds.), *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*, Washington, D.C.: National Academy Press.

Tucker, C., Fisher, S. K., Goldenberg, K. O'Brien, E. M., Rosen, R., Stinson, L. & Willimack, D. K. (2000, December), "Using Cognitive Methods to Improve Questionnaires for Establishment Surveys: A Demonstration." Paper presented at the meeting of the Federal Economic Statistics Advisory Committee, Washington, D.C.