REDESIGNING A QUESTIONNAIRE FOR COMPUTER-ASSISTED DATA COLLECTION: THE CURRENT POPULATION SURVEY EXPERIENCE

Cathryn Dippo and Anne Polivka
Bureau of Labor Statistics
2 Massachusetts Avenue, NE, Room 4915, Washington, DC 20212

Kathleen Creighton, Donna Kostanich, and Jennifer Rothgeb
Bureau of the Census
Washington, DC 20233

The conclusions in this paper are those of the authors and do not necessarily represent those of the Bureau of Labor Statistics or Bureau of the Census.

ABSTRACT

Over the last decade, there have been two new factors that have significantly influenced the design of survey data collection--the computer and the theories and methods of cognitive psychology. When, in 1986, staffs of the Bureau of Labor Statistics and the Bureau of the Census initiated a process for redesigning the Current Population Survey (CPS) for the 21st century, incorporating these two new factors was made a top priority. In this paper, we illustrate how, by concentrating on the cognitive processes of respondents and interviewers, we used computer-assisted interviewing as a tool for reducing measurement error.

Key words: cognitive aspects of survey methodology, questionnaire testing, measurement error

Acknowledgments: The new CPS questionnaire is the result of a team effort which involved many staff members from both BLS and Census. Space does not allow us to recognize everyone. The other members of the BLS-Census Questionnaire Design and Overlap Analysis Steering Committees over the years were Chester Bowie, John Bregger, Shail Butani, Lawrence Cahoon, Kennon Copeland, Harvey Hamel, Elizabeth Martin, Michael McMahon, Thomas Scopp, Clyde Tucker, Ronald Tucker, and Alan Tupek.

REDESIGNING A QUESTIONNAIRE FOR COMPUTER-ASSISTED DATA COLLECTION: THE CURRENT POPULATION SURVEY EXPERIENCE

Cathryn Dippo and Anne Polivka, Bureau of Labor Statistics
Kathleen Creighton, Donna Kostanich, and Jennifer Rothgeb, Bureau of the Census

1. Introduction

Over the last decade, there have been two new factors influencing survey questionnaire design--the computer and the theories and methods of cognitive psychology. Computers were first used by U. S. market research firms in the early 1970s to aid in the direct collection of data by telephone (Fink 1983). Since then, numerous computer-assisted telephone interviewing (CATI) systems have been developed. With the proliferation of light-weight portable computers over the last few years, the development of computer-assisted personal interviewing (CAPI) is proceeding rapidly. The theories of cognitive psychology were first explicitly used as the basis for redesigning the questionnaire for the National Crime Survey (Martin 1986). In the last few years, many of the major Federal statistical agencies and numerous private survey organizations have established "cognitive laboratories," which are routinely used in questionnaire development.

Our purpose in writing this paper is to promote the use of the above-mentioned new methodologies when designing a survey questionnaire by describing our experiences in designing an automated data collection instrument for the Current Population Survey (CPS). After providing some background material on questionnaire design and computer-assisted interviewing methodologies in section 2, we discuss the development of the CPS questionnaire over the last 50 years and how redesigning the CPS questionnaire for the 21st century has brought together the two new methodologies. In section 3, we highlight how we utilized the computer in evaluating

alternative questionnaire designs. Some examples are given in section 4 on the new CPS questionnaire's design features which aid the cognitive processes of the respondent and interviewer and are primarily dependent on the use of the computer. The effects of the new questionnaire and collection procedures on labor force estimates are briefly discussed in section 5. We conclude in section 6 with a discussion of issues for the future.

2. Background

The Current Population Survey is a monthly survey of approximately 60,000 households. The CPS survey, conducted for the Bureau of Labor Statistics (BLS) by the Bureau of the Census, is the primary source of information on the U. S. labor force. Each month BLS analyzes and publishes information from the CPS, such as the unemployment rate, demographic characteristics of individuals in the labor force, and the number of hours individuals work. The survey began in 1940 under the auspices of the Works Projects Administration and was called the Monthly Report of Unemployment. The current CPS questionnaire has remained essentially unchanged since the last major revision in January 1967. With only minor exceptions. the concepts measured have remained constant since the late 1940's.

Over its 50+-year history, the CPS has continued to be a model for survey designers. It was the first national probability sample of households, and many of the statistical methods for sampling and estimation now considered common practice were originally researched and implemented in CPS. When, in April 1986, the staffs of the Bureau of Labor Statistics and the Bureau of the Census met to discuss the future of the Current Population Survey, two of the six research areas identified as necessary to make CPS a leading-edge survey for the 21st century related to data collection--computer-assisted interviewing and the questionnaire. A Questionnaire Design Task Force was established to identify the cognitive and conceptual problems in the existing questionnaire, to suggest possible solutions for identified problems, and to develop a research plan to design and test a new questionnaire, along with related survey procedures. A separate task

force was established to investigate the potential uses of computer-assisted interviewing. When a final consolidated research plan was approved in 1988, a major premise of the plan was that all interviews would be conducted using a computer. Following a period of questionnaire development and extensive testing, Census began collecting au CPS data using a new fully-automated questionnaire in January 1994. CPS is the first monthly survey conducted by the Federal government to use fully-automated collection procedures.

In the rest of this section, we present background information on questionnaire design and computer-assisted interviewing methodologies and then discuss how we have blended innovations in these areas to produce a new questionnaire for CPS.

2.1. Questionnaire design methodology

Numerous texts and articles on the subject of questionnaire design have been published over the last 40 years. Some concentrate on question structure and sequencing. See, for example, *The Art of Asking Questions* (Payne 195 *1), Asking Questions: A Practical Guide to Questionnaire Design* (Sudman and Bradburn 1974), and *Survey Questions: Handcrafting the Standardized Questionnaire* (Converse and Presser 1986). Some concentrate on methods for testing and evaluating questionnaires. See, for example, *Approaches to Developing Questionnaires* (DeMaio 1983) and *TestinglAssessing Question Quality* (Thorslund and Wdrneryd 1985). Still others look at form design and other aspects of questionnaire administration.

While not necessarily explicitly stated, the over-arching concern in all these works is nonsampling error. The importance of acknowledging the existence of and measuring nonsampling errors has long been recognized (Hansen, Hurwitz, and Bershad 1961). Numerous studies have been undertaken to measure selected types of nonsampling error, including measurement error, response error, and interviewer variance. Groves (1989) describes the various methods used to estimate these errors.

However, studies designed to estimate measurement errors rarely provide insight into their causes, especially from a respondent's or interviewer's perspective. Our basic understanding of the underlying cognitive and social psychological processes used by respondents and interviewers during survey data collection is still minimal.

Although many of the authors of questionnaire design texts have had training in psychology and have investigated respondents' misunderstanding of survey questions, e.g., Belson (I981), the explicit call for an interdisciplinary approach to questionnaire development is relatively recent. Thomas (1979) addressed the question of why experimental psychologists and survey researchers thus far had so little that is useful and constructive to say to one another. The National Academy of Sciences, Committee on National Statistics, convened a panel on survey measurement of subjective phenomena. The panel included a cognitive psychologist and several other representatives from the social sciences. Their report (Tumer and Martin 1984) includes a recommendation for extensive interdisciplinary investigation of the subjective aspects of survey questions.

In 1983, the National Academy of Sciences sponsored a small conference on "Cognitive Aspects of Survey Methodology" (CASM), which was attended by psychologists, sociologists, anthropologists, and survey researchers. Their purpose was to investigate how the theories and research techniques of one discipline could benefit one another and thus build bridges between disciplines where none currently existed (Jabine, et al. 1984). Subsequently, the Social Science Research Council established a Committee on Cognition and Survey Research, which sponsored seminars on "The Effects of Theory-Based Schemas on Retrospective Data" and "The Role of the interviewer in Survey Measurement Error." The Committee also sponsored the book *Questions*

About Questions: Inquiries into the Cognitive Bases of Surveys (Tanur 1991) which provided an examination of the cognitive and social processes that influence the answers to questions.

There are two important aspects to the CASM movement relevant to a discussion of the questionnaire design methodology used for this redesign of the CPS questionnaire. One is the development of a psychological model that can relate psychological theory to how the questionnaire affects response and the other is the incorporation of laboratory techniques into the questionnaire design and testing process.

2. 1. 1. Psychological theory and questionnaire design

Many authors consider questionnaire design an art. For example, Platek (I985, p. 119) believes "Questionnaire development may be viewed both as a science and an art," while Converse and Presser (1986, p. 7) are less explicit when they discuss "strategies culled from examples or experience of question-crafters." One of the goals of the CASM movement has been to promote the development of questionnaire design as a science rather than an art. One aspect of this development process is the creation of a psychological model.

Tourangeau (1984) has presented a model of the respondent's task from a cognitive perspective which includes four steps: comprehension, retrieval, judgment, and response communication. Comprehension involves the assignment of meaning by the respondent to the survey question; retrieval--the memory search for relevant knowledge (concepts and facts); judgment--the selection of a response; and communication--the transmission of the selected response.

In discussing "The Current Status of Questionnaire Research," Bradbum and Sudman (I 99 1) categorize questionnaire response effects into question wording (understanding question words, words as cues, and the emotional content of words), question form (length of questions and specificity of the question, including open- versus closed-form), response categories (response categories as cues and scales, inclusion of *Don't know* response categories), and question flow (question ordering). How do we relate these factors which are known to affect response error to a psychological model? Let us take as an example question wording. The choice of words used in questions affect respondents' comprehension, because a word can be given no meaning by a respondent, a word can be taken to mean different things by a respondent, or a word can be taken to mean different things by different respondents (Groves 1989). The use of words as cues is directly related to respondents' retrieval of information from memory (Tanur 1991), and the emotional content of words may influence respondents' reporting through an effect referred to as "social desirability" (Groves 1989). The other factors historically known to affect respondent's behavior can be similarly related to the Tourangeau model.

As other behavioral scientists become involved in questionnaire design, specifically, and the data collection process, generally, other models are being developed. For example, Esposito and Jobe (1991) have presented "A General Model of the Survey Interaction Process," which advocates a more interdisciplinary view of the survey data collection process by incorporating theories from anthropology, psychology, sociology, and statistics. Mullin, Conrad, Sander, and Herrmann (1993) have developed more detailed cognitive models of the interviewer process of question generation and question clarification, the respondent process of question answering, and the interviewer-respondent interaction. These models should be viewed as frameworks for building a sound theoretical basis, so that questionnaire design can become a science rather than an art.

2.1.2. Laboratory methods and questionnaire design

Another aspect of converting questionnaire design from an art to a science is the development of a set of recognized investigative methods with criteria for evaluating their effectiveness in various situations. As discussed above with respect to psychological theory, this conversion means building upon a body of knowledge that has been accumulated over the years.

When the Subcommittee on Questionnaire Design of the Federal Committee on Statistical Methodology prepared their review of the methods used in Federal statistical agencies in developing questionnaires, they used terms such as *unstructured individual interviewing, qualitative group interview. informal testing, formal testing, and interviewing debriefing* (DeMaio 1983). When Forsyth and Lessler (1991) prepared a taxonomy of cognitive laboratory methods for the International Conference on Measurement Errors in Surveys, they used terms such as *expert evaluation, concurrent and retrospective think-aloud interviews, free-sort and dimensional-sort classifications, and focus group interviews.* Many of the procedures are basically the same in the two classification schemes, e.g., *qualitative group interview and focus group interview;* others differ because a more scientific hypotheses-based approach has been developed for the techniques among psychologists, e.g., Ericsson and Simon's (1980) study of verbal report methods. Fowler and Mangione (1990) and Fowler (1991) discuss principles and methods for designing a questionnaire to minimize interviewer effects. Methods for evaluating interviewer behavior have been developed over the last decade (Cannell, Lawson and Hausser 1975; Cannell and Oksenberg 1988). Reading questions, probing as a follow-up to inadequate responses, categorizing or recording verbatim responses to open-ended questions, and maintaining a professional, neutral relationship with the respondent are important aspects of the

interviewer's job which need to be taken into account when designing a standardized questionnaire.

2.2. Designing a questionnaire for computer-assisted interviewing

Despite the existence of CATI since the early 1970s, little was written on questionnaire design for computer-assisted interviewing until the mid- 1980s. And then, much of the literature focused on the questionnaire as a computer program (Nicholls and House 1987, Futterman 1988). Groves and Nicholls (1986) and House and Nicholls (1988) presented six basic objectives for CATI questionnaire design:

1. Collect survey data, including responses to questions, interviewer comments, and other relevant information.
2. Meet interviewer needs by having comprehensible displays, easy access to information, and methods for expediting movement within the questionnaire.
3. Ensure program correctness so the instrument functions as intended under all circumstances.
4. Function efficiently so that the CATI instrument makes minimum demands on hardware resources while maintaining rapid response times.
5. Provide documentation that ensures the program structure, coding, and operation are understandable to others.
6. Be portable to other surveys collecting similar information.

Each of these objectives were discussed in terms of software accuracy. Quality enhancements such as automatic branching and editing were most often promoted on the basis of controlling interviewer error. Little or no attention was given to exploiting the benefits of a computer as an

aid to the respondent's or interviewer's cognitive processes. Palit (1987), in his discussion of "Questionnaire Design for Automated Interviewing," stated that "In the movement away from a paper based system to a machine based system, I believe, the main danger is that we replace survey specialists with computer specialists. ... The data collection phase in the survey process is a human interaction problem and a subject matter problem first; only a very small part of the action can be claimed legitimately by the computer. We should not let the introduction of the computer into the collection process mislead us on this score."

Many of the computer-facilitated quality enhancements discussed by authors neglectful of the CASM perspective can be interpreted from a psychological modeling point of view. For example, question customizing using "fills" based upon prior answers such as the subject's name alleviates the need for the interviewer to comprehend that the..... in the question "Does... USUALLY work 35 hours or more a week at this job?" means she must substitute the subject's name, look at the roster of household members, make a judgment about which of the listed names should be substituted, and verbalize the new question. Similarly, automated branching and on-line help facilities can be viewed as procedures for reducing the interviewer's cognitive load.

Other enhancements cited by previous authors may, in fact, increase interviewers' cognitive load or have a negative effect on respondents' question-answering processes. For example, van Bastelaer, Kerssemakers, and Sikkel (1988) describe the advantages of having the computer detect inconsistencies in respondents' answers so that interviewers can reconcile them. Yet little thought and, as far as we know, no research has been expended to establish what are the best procedures for confronting respondents with what, in essence, amounts to making an error.

2.3. Designing a questionnaire for the Current Population Survey

Changes to the CPS questionnaire have always been approached very cautiously. The labor force questions used prior to January 1994 to measure the basic concepts of employment and unemployment are essentially the same as those used 25 years ago, when the last major questionnaire revision took place. The long time frame between major revisions is in part due to the need to gauge or benchmark the effects of the changes on labor force estimates--an expensive proposition.

Past experience with even a minor, grammatically innocuous change of a question from a double negative (Canamucio and Bushery 1986) indicates the sensitivity of the data collection process. Based upon the results of the Methods Development Survey, changing the question "Is there any reason why you could not have taken a job last week?" to "Could you have taken a job last week if one had been offered?'" was projected to have no effect on CPS estimates of unemployment. In the old question, a *no* response made the person available for work, hence, unemployed[1]. The revised question required a *yes* response for the individual to be considered available, hence, unemployed. What was not considered when the question was changed was the contaminating effect of the old question; it appeared that some interviewers continued to ask the old question by rote and then entered the response as no, resulting in the person being classified as not in the labor force rather than unemployed. (This was quickly resolved through additional training.)

The last major questionnaire changes to the CPS questionnaire occurred in January 1967. These changes, which were based upon recommendations of the President's Committee to Appraise

[1] Availability for work is one of the necessary conditions to be considered unemployed, along with active job search or being on layoff from a job to which an individual expects to be recalled.

Employment and Unemployment Statistics (known as the Gordon Committee) (U.S. Department of Commerce 1976), were preceded by several years of testing. Since then, the National Commission on Employment and Unemployment Statistics (the Levitan Commission), which was convened in the late 1970s, made a number of specific recommendations for changes to the CPS, some of which involved changes to the questionnaire wording and content. Although many of the recommendations were implemented (Butz and Plewes 1989), some were delayed until the recent revision in order to estimate the effects of the changes with a sufficiently large sample.

The 1988 final research plan for the CPS called for the development of a new questionnaire, which would be used in a fully-automated collection environment. That is, all interviewers, both in the field and in centralized telephone-interviewing facilities, would use a computerized questionnaire to collect data. Moreover, the plan called for the questionnaire to be developed using a CASM approach.

All of the changes to the questionnaire over the last 30 years have been constrained by the paper-based data capture system, which was predicated on the use of FOSDIC[2]. Some of the problems associated with this methodology are discussed in section 2.3.1 below. Thus, the decision to develop the new questionnaire within a fully-automated collection environment freed us from the physical constraints of a paper-based processing system developed many years before and allowed us to focus on developing a questionnaire which would reduce response and interviewer error. Some of the problems respondents and interviewers experienced with the old questionnaire are discussed in section 2.3.2.

[2]Film Optical Sensing Device for Input to Computer, one of the original optical character reading devices

2.3. 1. Physical and operational problems with old paper-based methodology

In the past, creative and efficient questionnaire design for the CPS was inhibited by four principal factors: limitations of the data capture system, space constraints, the physical separation of the household characteristics data collection form from the labor force questionnaire, and the inability to preprint information on the labor force questionnaire.

For over 30 years, FOSDIC was the primary data entry device for CPS. Despite its reliability and availability, FOSDIC technology had several liabilities that inhibited change to the core questions, limited the scope and design of supplement questions, and increased the possibility of interviewer error. The size of the CPS questionnaire booklet was fixed at 12 11-1/2-by-12 inch pages -- 4 pages for the core labor force questions, 4 pages for supplemental questions (when asked) and demographics, and 2 pages for capture of children, Armed Forces members, and household characteristics (in addition to outside covers). These restrictions made the core CPS questions, for the most part, immutable and limited supplement design flexibility.

With FOSDIC, information was captured by darkening a small preprinted circle that had to be within a prescribed distance to "index marks" in order to be read properly. Erasures were discouraged because they left enough of a mark to be read by FOSDIC. False "pickups" were also a possibility from stray markings on the questionnaire. Thus, interviewers had to take great care in completing and handling the questionnaire.

Another, much larger source of data entry error was inherent in the CPS data collection procedures whereby a "control card" which contained all the demographic background information for a household had to be kept within the interviewer's supervising office, while the FOSDIC-readable questionnaire was forwarded to a central processing facility for data capture and processing. This required that all the demographic and identifying information be transcribed by interviewers each month onto the FOSDIC-readable questionnaires before transmittal. This procedure was known to be error-prone in that errors of omission or inconsistency from month to month existed for these items. Moreover, while a full accounting of interviewers' assignments on a housing unit basis was undertaken clerically at the supervising office, there was no verification that each person listed on the control card had a labor force interview completed for it.

The procedures for the old CPS paper questionnaire made it operationally infeasible to provide any of the information recorded directly onto the FOSDIC-readable form back to the interviewers in succeeding months for use in dependent interviewing, that is, to read back to the respondent a previously reported answer to determine whether there had been a change. Yet, the responses to many of the core labor force questions remained the same from month to month for many people--for example, an employed person's job description from which industry and occupation codes were derived. The independent collection of this information resulted in over-estimates of month-to-month change (Collins 1975).

2.3.2 Respondents' and interviewers' cognitive problems with current questionnaire

Based upon the recommendations of the BLS-Census Questionnaire Design Task Force (Bregger, et al. 1986), which were expanded upon by Martin (1987), pre-questionnaire design work began with investigations of respondents' and interviewers' understanding of CPS concepts. Laboratory

research methods of paraphrasing, retrospective protocol analysis, and focus groups were used to investigate the key labor force questions used to determine whether individuals are employed, unemployed, or not in the labor force. The concepts examined included work, number of hours actually worked during a week, last week, looking for work. and being on layoff. Many of the laboratory techniques were used with individuals who had specific characteristics of interest. For example, focus groups were conducted with individuals who classified themselves as on layoff, looking for work, or as in households where at least two members worked. The majority of the research was conducted in cognitive laboratories in Washington, DC . However, a small proportion was conducted at sites where it was easier to recruit specific types of workers. For example, focus groups to assess individuals' understanding of what it means to be on layoff were conducted in a union hall in Baltimore, and focus groups to ascertain whether the measurement of marginal work activities differed among ethnic groups or occupations were conducted in Kansas City, KS, and Boulder, CO. For further details, see Palmisano (1989), Bienias (1988), and U. S. Bureau of Labor Statistics (1988).

To obtain a quantifiable measure of respondents' comprehension of labor force concepts, a respondent debriefing, consisting of special follow-up probes and hypothetical vignettes, was administered to 2,300 telephone respondents in CPS between July and December of 1988. The results of the respondent debriefing verified the findings of the laboratory research. Thus, for example, with respect to *last week*, when asked, only 58 percent of respondents reported that they thought the phrase referred to an exact time frame. In addition, when asked to define *last week*, only 17 percent of the respondents had a definition that matched the CPS definition of Sunday through Saturday. The majority of respondents, 54 percent, defined last week as Monday through Friday. Respondents' classification of vignettes indicated that the majority of respondents' concept of work was in accord with the CPS definition. However, a substantial minority had a broader definition, while an equally large number had a narrower definition of work than the CPS.

These variations in definition were revealed by the fact that 38 percent of the respondents classified as working a vignette situation that did not meet the official CPS definition, while, in another vignette situation, 50 percent failed to classify as working one that did meet the official CPS criterion. Corresponding to Schwarz's (1987) findings, analysis of the vignettes indicated that respondents' definitions of work depended on their characteristics and experiences. (For more information on this study, see Campanelli, Martin and Creighton (1989) and Martin, Campanelli, and Fay (1991).)

Interviewers' understanding of the CPS concepts were tested in March 1989, when they were administered a subset of the vignettes used in the respondent debriefing. The majority of interviewers classified the vignettes correctly. However, classification of activities as looking for work or marginal work were problematic for interviewers. Over a third of the interviewers indicated that activities which do not meet the official CPS definition, such as reading the newspaper, should be counted as legitimate job search methods (Campanelli, Martin, and Rothgeb 1991).

An overview of the old questionnaire's effectiveness from an interviewer's perspective was obtained through interviewer debriefings that consisted of a standardized questionnaire followed by focus group discussions (U. S. Bureau of Labor Statistics 1988). These interviewer debriefings were conducted with 120 interviewers at 13 sites across the country during March 1988. In these debriefings, interviewers indicated that, while on the whole the labor force questions worked effectively, there were several questions and segments they felt were subject to response error. These included the underreporting of work activities, especially irregular, illegal, marginal, or non-monetarily rewarded work, the overreporting of job search activities due to

respondents' desire to provide socially acceptable answers, and a lack of a common understanding of phrases such as *last week.* Only 52 percent of the interviewers believed that all or nearly all respondents' understood the CPS definition of work. Seventy-four percent of the interviewers felt that respondents some of the time or most of the time said they were looking for work because they felt it was expected that they should be working or looking for work.

Taken as a whole, the examinations of respondents' and interviewers' comprehension indicated that concepts such as *last week and layoff* needed to be made explicit in the questionnaire. A method for distinguishing between active[3] and passive job search methods needed to be incorporated into the questionnaire, and the structure of the questionnaire, along with question wording, needed to be altered to broaden the measurement of work activities.

At the same time that interviewers' and respondents' comprehension were being analyzed, other aspects of their mental processing were being investigated. For example, an extensive analysis was conducted of interviewers' judgments in the coding of verbal responses to open-ended questions into pre-specified categories. This analysis involved several steps. First, the frequency distributions of responses for each of the open-ended CPS questions were analyzed to identify infrequently used response categories and over use of the *other* category. Subsequent to the response distribution analysis, a sample of responses were analyzed to formulate new categories. These sample responses were obtained from three sources: CPS telephone interviews conducted from the centralized CATI facility between July and December 1988 (7,200 transcribed *other* responses distributed over six questions); CPS field interviews (personal visit or telephone) conducted between November 1988 and February 1989 (981 responses to two questions asking

[3] An active job search method is one which has the potential to result in a job offer from an employer without further action by the job seeker. For example, sending a resume to an employer in response to reading a want ad is ad active job search activity, while just reading the want ads is a passive job search method.

the reason for working less than 35 hours); and the CPS questions in the National Longitudinal Survey interviews conducted in 1988 (at least 1,000 responses to each question). New categories were formulated using a q-sort methodology. The accuracy and consistency of interviewers' coding using the new response categories were then evaluated using a subset of the verbatim responses and a sample of interviewers. For a complete description of this research project, see Fracasso (1989). Other projects which used laboratory methods were reported on by Boehm (1989); Edwards, Levine, and Allen (1989); and Gaertner, Cantor, Gay, and Shank (1989).

Throughout the preliminary research, the focus was on the cognitive aspects of respondents' and interviewers' tasks. The computer was used as a tool for designing a question and answer process which eased the cognitive processing of the participants. Many of the characteristics of the questionnaire which are facilitated by using a computer are discussed in section 4, while the use of the computer as a tool for evaluating alternative questionnaires is discussed in section 3 below.

3. Utilizing the computer in questionnaire evaluation

Based on the results of the preliminary research, recommendations of the joint BLS-Census Questionnaire Design Task Force, and the National Commission on Employment and Unemployment Statistics, two alternative versions of the questionnaire were developed. These two alternatives were compared to the current questionnaire in a large computer-assisted telephone interviewing test using a random-digit dialing sampling procedure (CATI/RDD). This test was conducted from July 1990 to January 1991 and included interviews for about 72,000 individuals. The primary objectives of this CATI/RDD test were to provide data to aid in the selection of the best of the alternative wordings for each question (or question set) and to identify problems in question wording or sequencing.

Based on the results of the first CATI/RDD test, a third alternative version of the questionnaire was developed. Like the first two alternatives, the third alternative questionnaire was tested against the current questionnaire in a CATI/RDD test. The second CATI/RDD test was conducted from July 1991 to October 1991 and consisted of interviews for approximately 30,000 individuals. The purpose of the second CATI/RDD test was to obtain data for finalizing question wording and sequencing for the new questionnaire.

Both of the CATI/RDD tests essentially involved moving the cognitive laboratory to the field. The data from the CATI/RDD tests were primarily analyzed using four methods-systematic behavior coding of interviewers' and respondents' interactions, interviewer debriefings, field-based respondent debriefings, and item-based response analysis. The use of the computer greatly enhanced, increased the efficiency, or made feasible three of these four analytical methods. Moreover, the computerized instrument made it both possible and cost-effective to use all these techniques simultaneously. (For further information on the methods, as well as a discussion of their strengths and weakness, see Campanelli, *et al* . 1991 and Esposito, *et al* 1991.)

3. 1. Behavior coding

Behavior coding consists of quantifying the number of questions interviewers read incorrectly or respondents have difficulty answering. It can be done in real-time by monitoring interviews in progress or by using tape-recorded interviews. Behavior coding data are useful in identifying problematic items in a questionnaire. Comparing behavior coding data across questionnaires is one way to determine if alternative questions improve the quality of interviewer-respondent

interactions. That is, if one version of a question results in more accurate reading by interviewers and more adequate responses by respondents, it may be considered better.

All of the interviews for the CATI/RDD test were conducted from the Bureau of the Census's CATI facility in Hagerstown, MD, where it was possible to monitor interviews in progress. Using a specially developed form inspired by the work of Cannell et al. (1989), six researchers from BLS and Census coded interviewer-respondent exchanges, noting whether interviewers read a question exactly as worded, with a slight change in wording, or with a major change. Monitors also noted whether respondents gave adequate answers, qualified answers, inadequate answers, asked for clarification, interrupted the questions, said they did not know the answer to the question, or refused to answer. By monitoring interviews conducted from the CATI facility, it was possible to behavior code large numbers of geographical disperse interviews in a single day.

Computer-assisted interviewing at the centralized facility also made it possible for the monitors to simultaneously see and hear the questions interviewers were asking, enhancing their ability to code an interview while it was in progress. Otherwise, given the complicated skip patterns and the numerous fills in the revised CPS questionnaire, it would have been extremely difficult to determine if the interviewer was asking the question exactly as worded. Monitoring of question reading could have been done with taped interviews, but having the question appear on the screen reduced the probability that the appropriate fill or question wording would not be apparent to the monitor.

Another advantage of monitoring from a centralized computer-assisted interviewing facility is that interviews could be monitored more anonymously. Requesting interviewers to tum on tape

recorders would have been far more intrusive. The more discrete nature of monitoring in a computer-assisted interviewing facility, hopefully, would reduce the bias introduced by interviewers being aware that they were being monitored.

3.2. Respondent debriefing

Respondent debriefing involves incorporating follow-up questions in a field test interview to gain a better understanding of how respondents interpret questions asked of them. Respondent debriefings can be used to gather information similar to that acquired by some laboratory methods. However, by conducting respondent debriefings in conjunction with the CATI/RDD test, it was possible to draw inferences for a population much larger than is possible with laboratory methods.

Computer-assisted interviewing allowed the CATI/RDD respondent debriefings to differ in several ways from debriefings typically administered using paper instruments. First, follow-up probes were tailored for each household based upon responses to the main survey. For example, individuals indicating they had more than one job were asked how they determined which job was their main job. Second, computer-assisted interviewing also made it possible to prioritize different segments of the respondent debriefing. A high priority was assigned to follow-up probes for rare or hard-to-find subpopulations. For example, since unpaid family workers constitute a very small proportion of the American work force, persons in households reporting a business or farm and who did not report working during the reference week were asked questions to ascertain whether they had done any unpaid work. The priorities of the follow-up probes allowed a wide variety of debriefing questions to be asked without overburdening respondents or jeopardizing reliability.

For those households not belonging to one of the targeted subpopulations--the majority, a sampling algorithm was used to determine which of several sets of debriefing questions a household would receive. In addition, computer-assisted interviewing made it possible to use a fractional replication factorial design (Alexandria and Becker 1978) to modify questions which might be subject to gender bias and randomize question order (Perreault 1975) to avoid context effects.

3.3. Response analysis

The primary purpose of analyzing response distribution data is to determine if the different question versions or different question sequencing produce differences in response patterns that affect labor force estimates. Response distribution data were analyzed for all questionnaire items.

The analysis of nonresponse and response distributions was facilitated by the computer in several ways. First, data were available quickly. (Due to the tightness of the overall schedule, data from the first CATI/RDD test conducted between July 1990 and January 1991 needed to be analyzed and incorporated into a revised alternative questionnaire by April 1, 199 1. Data from the second CATI/RDD test conducted from July 1991 to October 1991 needed to be analyzed and incorporated into a final revised questionnaire by January 1, 1992.) By eliminating the need to translate data from a paper instrument into computer-readable form and by including some preliminary data editing and classification recoding into the computerized instrument, the time required for data processing was reduced.

Second, verbatim responses to open-ended questions were available simultaneously with interviewers' coding of the responses into prespecified categories. Currently in the CPS, recorded verbatim answers are stored on microfiche, making access difficult and time-consuming. Having the transcribed answers in a computer-readable form correctly associated with the marked response category, made it possible to analyze quickly whether the precoded response categories were being used properly and to determine if additional or different response categories were necessary.

The use of the computer also made it possible to collect nonresponse data in a more useful and probably less biased way than if a paper instrument had been used. On a paper questionnaire, answers of "don't know" or refusals typically are handled in one of three ways--by including *don't know and refusal* response categories, by instructing interviewers to leave the answer blank, or by interviewers obtaining a "best guess" for an answer. Each of these approaches has its disadvantages. When *don't knows and refusals* appear as printed response categories, they serve as cues or reminders to interviewers that these are acceptable responses. It is suspected that these cues increase the use of these categories and thus may bias the results. If, on the other hand, the interviewer is instructed to leave answers blank, it is impossible to distinguish situations where the interviewer accidentally left an answer blank from instances where respondents did not provide an answer. In addition, it is impossible to separate *don't knows from refusals.* If. using the third procedure, interviewers are instructed to obtain an answer even if it entails guessing, the quality of the data may be compromised, since answers which were guesses or elicited after extensive prodding usually are not identified.

With a computerized survey instrument, it is possible to obtain a "truer," more usable measure of nonresponse by including *don't know and refusal* as response categories, but not having them appear on the interviewer's screen. These data made it possible to select questions with the lowest overall nonresponse rate and to distinguish questions requiring a difficult cognitive task from sensitive questions through the analysis of don't know rates versus refusal rates.

4. Design features of computer-assisted instruments

An instrument designed for computer-assisted interviewing should be considerably different from one designed to be printed on paper. Computerized data collection instruments should be thought of as expert systems or as software tools with built-in knowledge bases that aid users in performing complex cognitive tasks. When considered from this perspective, the design of a computer-assisted instrument has two major components: the development of a knowledge base and the description of the cognitive task. The knowledge base for CPS, for example, includes the detailed working definitions for each labor force category, the detailed coding and classification systems for industry and occupation, etc. An example of a cognitive task model is given by Mullin, et al. (I 993), which includes a break down of both interviewers' and respondents' jobs into comprehension, recall, judgment, and verbalization/recording tasks. The designer's job is to develop a tool which aids interviewers and respondents in their cognitive tasks by taking advantage of the computer's ability to store information and perform complex functions. For example, question tailoring and complex question sequences can be used to aid respondents in comprehending what information is being sought; standardized probes can be built into the instrument to aid the interviewer in recalling all the bits of information needed for classification; information can be brought forward from previous interviews (dependent interviewing) to reduce respondent recall error; and response categories for open-ended questions can be viewed as cues to aid interviewer's judgment.

In this redesign of the CPS questionnaire, we have not completely incorporated the knowledge base associated with CPS. That is, we have attempted to build into the instrument questions and response categories which incorporate the basic knowledge needed to classify individuals into a major labor force category. We have not incorporated an on-line help system containing the detailed information in the interviewer's manual pertaining to special situations. And, while plans exist for incorporating automated industry and occupation coding into the instrument, this would only be a first step toward making the CPS an expert system for associating an accurate industry and occupation code to each interviewed individual,

We have, in this redesign, focused on the cognitive tasks of the interviewer and respondent. The new CPS questionnaire was designed to reduce measurement error and respondent burden, while permitting the collection of more information and incorporating validity checks within the interview. For the sake of brevity, we will only discuss selected design features which facilitate the cognitive tasks of the interviewer and respondent-.complex question sequencing, question tailoring, standardized probes, dependent interviewing, and built-in editing procedures.

4. 1. Complex question sequencing

Computer-assisted interviewing permits complicated question-to-question branching instructions or skip patterns to be incorporated into the questionnaire without adding the potential for interviewer error. Computerized skip patterns can use information from several questions simultaneously.

4.2. Question tailoring

The tailoring of questions to reflect answers to preceding questions aids respondents in comprehending what information is being sought. For example, during the first CATI/RDD test of the question, "Last week did you do any work for pay or profit?," interviewers reported that a response sometimes heard was "No, I didn't work for profit, but I worked at my regular job." This response indicated that the respondent missed the word *pay* in the question and only heard the word *profit.* To avoid this misunderstanding, the question was modified to reflect the answer to a previous question on the existence of a family business. Specifically, if it had been previously indicated that there was a business within the household, the question would include the phrase "either for pay or profit." For households in which no business was indicated, the question would simply be "Last week, did you do any work for pay?"

Question tailoring can also reduce respondent burden. For example, one of the most common complaints about the old CPS questionnaire was that it was burdensome for retired persons who had no attachment to the labor force (BLS 1988). Respondents who indicated they were retired in response to the question inquiring about their major activities last week were still asked if they worked last week, if they were temporarily absent from a job or on layoff, if they had looked for work in the last 4 weeks, and, in their fourth and eighth interviews, their work histories within the last 5 years. In the redesigned survey, the response category "retired" is included for each question inquiring about an individual's labor force status. If individuals volunteer that they are retired in response to any of these questions, the computer verifies that they are 50 years or older and skips these respondents to questions inquiring whether they want a job and when they last worked. If they indicated that they do not want a job, the individuals are classified as retired and the interview is terminated.

4.3. Standardized probes

The incorporation of standardized probes into the CPS instrument should reduce errors caused by variation in interviewing procedures. For example, in coding an individual's industry and occupation, it is important to know if the individual's business or industry is mainly manufacturing, retail trade or wholesale trade. In the old questionnaire, respondents were first asked for whom they worked and then "What kind of business or industry is this?" There was no guarantee that the respondent would volunteer the information that the business was manufacturing. retail trade, wholesale trade, or some other industry. Nor was there any guarantee that the interviewer would remember to probe for this information. In the new computerized instrument, the question inquiring about the kind of business or industry in which someone is employed is followed by an interviewer instruction to ask if necessary the question "Is this business or organization mainly manufacturing, retail trade, wholesale trade, or something else?"

In addition to insuring uniformity among interviewers, the incorporation of standardized probes into the survey instrument should also reduce measurement error by providing a method to check respondents' understanding of questions; by addressing the needs of selected subgroups of respondents; and by encouraging more precision in respondents' answers. For example, standardized probes were added to elicit information on job search, clarify the kind of information needed about why an individual worked part time, and determine if an individual whose usual hours vary usually works full time.

An example of the use of probes to permit respondents to answer in a more natural way, while obtaining more precise information, occurs in the questions about how long someone has been

looking for work or has been on layoff. Analysis of the first CATI/RDD test data confirmed the findings of Bowers and Horvath (1984) that even when respondents report the length of time they have been unemployed in weeks, they have a tendency to report the time in 4 week or month long intervals. Bowers and Horvath also found that there was a tendency for respondents to overreport the length of unemployment for shorter unemployment spells, but underreport the length of unemployment for longer spells. To encourage more precision in reporting without forcing respondents to report the data in an uncomfortable periodicity, respondents can now report their duration in weeks or months. However, if they report a duration of 4 months or less, a standardized probe is administered to obtain the duration in weeks.

4.4. Dependent interviewing

Prior to the redesign, all labor force information was collected independently each month. For information which does not change frequently, the re-collection of this information is inefficient, burdensome and a source of irritation for respondents. For example, in monitoring interviews, when respondents came to the questions inquiring about their places of employment and usual activities or duties, a frequent response was "the same as last month" or "I told you that last month." Furthermore, the 1975 Collins study and the results of the respondent debriefing in the CATI/RDD tests indicated that a substantial proportion of the monthly changes in industry and occupational codes was spurious. Specifically, the CATI/RDD tests indicated 38.8% of those reporting for themselves who indicated that neither their place of employment or theirjobs duties changed, had different occupation codes in consecutive months, while 21.4% of those who said they had not changed employers had different industry codes. On the other hand, because of the tendency for dependent interviewing to result in underestimates of change, it was implemented in the new, computerized questionnaire in only three areas--confirmation of retired status, industry and occupation, and length of time looking for work or on layoff.

Individuals employed in consecutive months are read back the name of the company for whom they were reported working in the previous month and asked if they still work there. If they confirm that where they worked last week is the same as previously recorded, they are asked if their usual duties and activities have changed since last month. If respondents indicate that there have been no changes, they are read back the description of duties and activities recorded in the previous month and asked to verify their accuracy. The results of the second CATI/RDD test indicated that approximately 5 percent of respondents changed 3-digit industry month to month, while 7 percent changed 3-digit occupations[4]. These figures were much closer to estimates of true change obtained by Westat under contract to BLS of 3.8 to 4.2 percent change between 3-digit industry categories and 5.9 to 7.4 percent change between 3-digit occupations.

Consistency in the reporting of unemployment duration was also improved by using dependent interviewing. In the first CATI/RDD test, it was found that only 24.8% of those looking for work in consecutive months increased their reported time of unemployment by 4 weeks plus or minus a week[5]. The revised CPS instrument eliminates this inconsistency by automatically adding the appropriate number of weeks to the initially reported duration for persons who have been unemployed in consecutive months.

[4] These estimates of change include month-to-month change for both dependent and independent measures of industry and occupation.

[5] Bowers and Horvath (1984) estimated that in 1976 only 24.4% of those unemployed in consecutive months reported durations that were consistent, where consistency was defined as the second reported duration being 4 or 5 weeks more than the first duration.

4.5. Built-in editing procedures

When one views a computer-assisted data collection instrument as an expert system, it seems natural to include as much as possible of what is typically viewed as *post-collection processing* into the instrument. This includes, for example, data verification procedures and range and consistency edits. By detecting possible errors and resolving inconsistencies during the interview when respondents can aid in the correction of the data, the quality of the data should be improved. Without these procedures, inconsistent answers can only be "corrected" through imputation in post-interviewing processing. In the redesigned CPS, three types of verification procedures were used--verification by respondents that their answers were recorded correctly, verification by respondents of calculations done by the computer, and verification by interviewers that they correctly entered respondents answers.

Examples of each of these verification procedures can be found in the redesigned earnings questions. Within the earnings series, respondents are encouraged to report their earnings for the time period with which they are most comfortable, e.g., hourly or annually. Respondents who have chosen to report their earnings for a time period other than hourly are asked the following question for verification purposes. "I have recorded your total earnings as (recorded amount) (WEEKLY, BFWEEKLY, MONTHLY, or ANNUALLY) before taxes or other deductions. Is that correct?" If the respondent indicates that the answer is incorrect, the earnings question is repeated ONCE using the time frame for which the respondent expressed a preference. The respondent is not asked to verify the second answer, since a second verification could be considered badgering by the respondent and place the interviewers in an awkward situation.

Persons who choose to report their earnings hourly are also asked to report how many hours they usually work at that rate of pay. Further, if these individuals indicate that they receive overtime pay, tips, or commissions, they are asked to report these earnings separately. The computer than calculates usual weekly earnings, and respondents are asked to verify the computer's calculations through the question, "I have estimated your usual WEEKLY earnings as (computer calculated amount) before taxes or other deductions. Does that sound correct?" If a respondent indicates that the calculated usual weekly earnings does not sound correct, a question is asked to determine which piece of information is incorrect. When the respondent identifies the incorrect piece or pieces of information, the question or questions used to collect the information are repeated. The computer then recalculates the usual weekly earnings, and the respondent is asked to verify the second calculation. If the respondent says the calculation still does not seem correct, the interviewer indicates that the respondent still thinks the answer is incorrect. The irreconcilable difference is logged within the instrument, and the interview continues with the next series of questions in order to avoid badgering the respondent or placing the interviewer in an awkward position.

The third type of verification in the redesign--interviewer verification of data entry--also occurs in the earnings series. If a recorded answer falls outside of a prespecified range, the following message appears on the screen

DO NOT READ

Weekly earnings recorded as (fill) weekly[6]

IS THIS ENTRY CORRECT?

[6] If the respondent is reporting hourly, biweekly, monthly, or annually, the periodicity which appears on the screen is changed accordingly.

Interviewers receiving an indication that an error in recording the answer has been made are asked to reenter the answer. If the computer does not indicate any typing errors, the interview continues to the next series of questions.

5. Effects of redesign on major labor force estimates

For 18 months prior to incorporating the redesigned instrument and collection methods into the CPS in January 1994, they were tested in a parallel survey. While designed to provide annual average labor force estimates for the total United States which could be compared to similar estimates from CPS to assess the overall effect of the redesign, the Parallel Survey contained subpanels which could be used to estimate the effects of specific combinations of centralization[7], computer-assisted interviewing, and questionnaire changes. Since the total sample for the Parallel Survey (PS) was relatively small (12,000 households per month, as compared to 60,000 in CPS), the size of the subpanels was in many cases as little as 450 households per month. While we recognized from the onset that the possibility was small of finding statistical evidence that some of the redesigned procedures were affecting labor force estimates, we felt it was important to design the Parallel Survey to obtain as much information as possible.

5. 1. Overall effect of changes in the questionnaire and collection methodology

The overall estimated annual unemployment rate as measured in the Parallel Survey for 1993 was 0.5 percentage point higher than the overall estimated annual unemployment rate as measured by CPS. The difference between the estimated unemployment rates for the two surveys was

[7] Data collection from the two Census centralized CATI facilities in Hagerstown, MD, and Tucson, AZ.

significant for women (adult women, white women, black women, other race women and Hispanic women), teenagers, and those 65 years and older. Moreover, the effect of the redesign changes was differential by sex, with the percentage increase in women's unemployment rate (10.2 percent) being significantly higher than the percentage increase in men's unemployment rate (4.0 percent), although there was not sufficient evidence to conclude there was an increase for men.

While the estimate of the overall employment-to-population ratio was not affected by the redesign, the redesign had a differential effect by gender. The estimated ratio for men was significantly lower in the Parallel Survey (69.3 percent) than in CPS (69.9 percent), while the estimated ratio for women was significantly higher (54.9 percent in the Parallel Survey compared to 54.2 percent in CPS).

Given these important differences, detailed analysis were conducted, comparing estimates from the two surveys (Polivka 1994) and assessing the effects of the redesign on data quality (Rothgeb 1994). Moreover, extraordinary efforts were put into developing estimates which would help users bridge the gap between the two surveys. These efforts included providing monthly estimates for 1993 from the Parallel Survey using a measurement error model (Miller 1994), predicting monthly estimates for early 1994 of what the old procedures would have measured (Tiller and Welch 1994), and assessing whether other design differences could be contributing to the estimated 0.5 percentage point difference in the unemployment rate (Kostanich and Cahoon 1994).

5.2. Effects by selected methodological change

Using subpanels embedded in the CPS and Parallel Survey, we were able to conduct an analysis of four potential data collection effects--combined questionnaire and computer-assisted interviewing, combined centralization and computer-assisted interviewing, centralization, and questionnaire. Considerably larger sample sizes were available for testing the two combined effects, whereas the sample sizes for the two singular effects were much smaller. Consequently, we would only be able to detect extremely large centralization or questionnaire effects. Thus, in this respect, the results from each of the tested hypothesis are not comparable. Other differences in the design of the subpanels further limits the comparability of the tested effects. See Thompson (1994) for a more extensive discussion of the designs and limitations of this analysis.

We found evidence to reject the hypothesis that the combined effect of changing the questionnaire and using computer-assisted interviewing had no effect on estimates of the unemployment rate (overall and for various demographic subpopulations) and women's (total and white) labor force participation rate. Tests of this hypothesis were based on month-in-sample panels I and 5 from the Parallel Survey and CPS, when almost all interviews are conducted by Census field representatives making personal visits to sample households and there is no telephone interviewing from the centralized CATI facilities.

We also found evidence to reject the hypothesis that the combined effect of computer-assisted interviewing and centralization had no effect on estimates of the unemployment rate (overall and for white men and women and black women). However, we found no evidence of computer-assisted interviewing combined with centralization on civilian labor force participation rate. Tests

of this hypothesis were based on subpanels created from CPS sample in purposely selected large metropolitan areas. Within these areas, statistically representative split panels were formed such that sample in one of the panels was eligible for a computer-assisted interview from a centralized facility and the remaining sample was interviewed by field representatives using the paper-and-pencil questionnaire. The old questionnaire was administered to both panels, although the computer version was a slight modification of the paper-and-pencil version. See Shoemaker (1993) and Thompson (1994) for more details.

Tests of the hypotheses related to a centralization effect and a questionnaire effect were less conclusive. In both cases, only the estimates of the unemployment rate for all blacks were found to be significantly different[8] at a lower level of confidence. However, as noted above, these tests were based on extremely small sample sizes and can only detect very large differences. Tests of the centralization hypothesis were based on subpanels created from the Parallel Survey sample in all areas that had a multiple number of interviewers. Split panels were formed in a manner similar to CPS with sample in one panel eligible for centralized interviewing. Only one-tenth of the sample in these areas was not eligible for centralized interviewing. The new questionnaire was administered in both panels. Tests of the questionnaire hypothesis were based on panels in the CPS and Parallel Survey that were eligible for centralized interviewing and were representing the same geographic areas.

This analysis presents convincing evidence that data collection method does effect estimates, particularly the two combined effects of centralization plus computer-assisted interviewing and questionnaire plus computer-assisted interviewing. We cannot, however, determine how each of

[8] The black male rate was significant for the centralization effect, while the black female rate was significant for the questionnaire effect.

these data collection methods is contributing to the overall redesign effect, nor can we determine which method of data collection is having the strongest effect.

6. Issues for the future

As stated in section 2.3, the goals for this redesign of the CPS questionnaire were to improve data quality by facilitating the cognitive processes used by interviewers and respondents. We concentrated on issues of word and concept comprehension. To a lesser extent, we attempted to reduce errors in recall by building specific recall strategies into the questioning process and to reduce errors in judgment by reducing the reliance on respondent's volunteering of information.

We in no way attempted to tackle all the factors which may affect the cognitive processes of respondents and interviewers. Even if we ignore behavioral, personality, and learning issues related to interviewer selection, interviewer training, respondents' decision to participate or not in the survey, etc., and limit the discussion of issues for the future to the cognitive issues and the questionnaire, the list is long. Therefore, we present just a few to aid the reader's understanding of the breadth of issues needed to be considered when designing a questionnaire for computer-assisted interviewing.

The use of a computer for displaying questions and entering responses instead of paper and pencil brings about a different set of "form design" concerns. Video intensity, character size, video display type and angle, keyboard layout, and weight (for CAPI) are factors which may affect interviewer behavior (Couper, Groves, and Jacobs 1990). Compared to what is known about the effects of form design parameters, such as white space, font type, and paper or print color, on the

mental processing of respondents using self-administered questionnaires, little is know about the effects of screen layout on interviewer's mental processing.

Some of the hardware and software issues often cited with respect to objective 4 (see section 2.2) are also relevant when considering the questionnaire as an aid to the mental processing of respondents and interviewers. For example, screen delay and slow computer response times may promote discomfort or anxiety in interviewers and thus affect their behavior. They might, for example, resort to question paraphrasing or recall. (Nervousness and anxiety have been shown to affect memory processing (Eysenck 1982, Reason 1988).) Evidence of the possible effects were found in CATI and CAPI tests for the National Health Interview Survey. More than half of the interviewers involved in a CATI test cited response time as a problem (Groves and Mathiowetz 1984), and interviewers involved in a CAPI test tended to enter briefer comments than when using paper and pencil because of a perceived need to reduce the "dead" time between questions (Rice, Wright and Rowe 1988).

Groves, Berry, and Mathiowetz (1980) discussed the need for research on segmentation or the inability of interviewers to examine or access the questionnaire as a whole. Despite their call for research over 10 years ago, there are no known studies to determine if contextual ignorance exists when a physical questionnaire does not exist, much less what the effects are.

Nor are all the issues for the future related to the interface of humans with computers. There is much to be done to convert the CPS into an expert system containing all of the knowledge currently contained in the interviewer's manual and training guide, coding manuals, and other paper documents. Moreover, even in the area of question wording where we concentrated our

efforts, we feel that we were particularly unsuccessful with one set of questions--the identification of industry and occupation. The problem of accurate classification is not simply a coding problem, nor is it a simple question design problem. The complexities of the current Standard Industrial Classification and Standard Occupational Classification systems are such that the information needed for accurate coding varies, making it impossible to devise 4 or 5 standard questions to collect the data necessary to categorize everyone accurately. In addition, an investigation of the translation of the verbal responses by respondents into industry and occupational codes indicated inconsistencies in the conversion of the verbal responses into written descriptions by the interviewers sometimes resulted in different codes. Thus, one issue for the future is the development of a true expert system within the CPS instrument for associating an accurate, detailed industry and occupation code to each employed person.

7. Summary

The data produced from the CPS are closely-watched by economic forecasters and policy analysts. Therefore, all changes need to be carefully researched prior to implementation. By concentrating on facilitating the cognitive processes used by respondents and interviewers, research on alternative measurement processes can result in reduced nonsampling errors. By capitalizing on the power and versatility of the computer, new research tools can be developed to provide the evidence needed to understand the effects of changes in data collection procedures. We hope that the approach we have used for developing the new measurement process for CPS will serve as a model for future survey redesign projects.

References

Alexander, C.S. and Becker, H.J. (1978). The Use of Vignettes in Survey Research. Public Opinion Quarterly, 42, 93-104.

Belson, W. A. (I 981). The Design and Understanding of Survey Questions. London: Gower.

Bienias, J. (1988). Preliminary Laboratory Research on Current Population Survey Questionnaires with Employed Persons. Proceedings of the Section on Survey Research Methods, American Statistical Association, 185-190.

Boehm L.E. (1989). Reliability of Proxy Response in the Current Population Survey. Proceedings of the Section on Survey Research Methods, American Statistical Association, 486-489.

Bowers, N. and Horvath, F. (1984). Keeping Time: An Analysis of Errors in Measurement of Unemployment Duration. Joumal of Business and Economic Statistics, 2, 140-49.

Bradbum, M. and Sudman, S. (1991). The Current Status of Questionnaire Research. In P. Biemer, et al. (eds.), Measurement Effors in Surveys, New York: Wiley.

Bregger, J., et al. (1986). Report of the BLS-Census Bureau Questionnaire Design Task Force. Staff report dated November, Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics.

Butz, W.P. and Plewes, T.J. (1989). A Current Population Survey for the 21st Century. Proceedings of the Fifth Annual Research Conference, Bureau of the Census, 3-13.

Campanelli. P.C., Martin, E.A., and Creighton, K. P. (1989). Respondents' Understanding of Labor Force Concepts: Insights from Debriefing Studies. Proceedings of the Fifth Annual Research Conference, Bureau of the Census, 361-374.

Campanelli, P.C., Martin, E.A., and Rothgeb, J.M. (1991). The Use of Respondent and Interviewer Debriefing Studies as a Way to Study Response Error in Survey Data. The Statistician, 40, 253-264.

Canamucio, A. and Bushery, J.M. (1986). The Rewording of Item 22E and Its Effects on CPS Data. Internal Memorandum Dated June 12, Washington, DC: U.S. Department of Commerce, Bureau of the Census.

Cannell, C.F., Lawson, S.A., and Hausser, D.L. (1975). A Technique for Evaluating Interviewer Performance. Ann Arbor, MI: The University of Michigan.

Cannell, C.F., and Oksenberg, L. (I 988). Observation of Behavior in Telephone Interviews. In R.M. Groves, et al. (eds.), Telephone Survey Methodology, New York: Wiley.

Cannell, C., Kalton, G., Oksenberg, L. and Bischoping, K. (1989). New Techniques for Pretesting Survey Questions. Final Report for Grant No. HS 05616 from the National Center for Health Services Research and Health Care Technology Assessment, Ann Arbor, MI: Survey Research Center, University of Michigan.

Collins, C. (1975). Comparison of Month-to-Month Changes in Industry and Occupation Codes with Respondent's Report of Change: CPS Job Mobility Study. Unpublished Bureau of the Census Response Research Staff Report No. 75-5 dated May 15.

Converse, J. M. and Presser, S. (1986). Survey Questions: Handcrafting the Standardized Questionnaire. Beverly Hills, CA: Sage Publications.

Couper, M., Groves, R.M., and Jacobs, C.A. (1990). Building Predictive Models of CAPI Acceptance in a Field Interviewing Staff. Proceedings of the 1990 Annual Research Conference, Bureau of the Census, 685-702.

DeMaio, T. J., ed. (1983). Approaches to Developing Questionnaires. Statistical Policy Working Paper 10, U.S. Office of Management and Budget, Office of Information and Regulatory Affairs, Washington, DC: Government Printing Office.

Edwards, W.S., Levine, R., and Cohany, S. (1989). Procedures for Validating Reports of Hours Worked and for Classifying Discrepancies Between Questionnaire Reports and Validation Totals. Proceedings of the Section on Survey Research Methods, American Statistical Association, 496-501.

Ericsson, K.A. and Simon, H.A. (1980). Verbal Reports as Data. Psychological Review, 87, 215-251.

Esposito, J.L. and Jobe, J.B. (1991). A General Model of the Survey Interaction Process. Proceedings of the 1991 Annual Research Conference, Bureau of the Census, 537-560.

Eysenck, M.W. (1982). Attention and Arousal: Cognition and Performance. Heidelberg: Springer-Verlag.

Fink, J.C. (I 983). CATI's First Decade: The Chilton Experience. Sociological Methods & Research, 12, 153-168.

Forsyth. B. and Lessler, J. (1991). Cognitive Laboratory Methods: A Taxonomy. In P. Biemer, et al. (eds.), Measurement Errors in Surveys, New York: Wiley.

Fowler, F.J. (1991). Reducing Interviewer-Related Error Through Interviewer Training, Supervision, and Other Means. In P. Biemer, et al. (eds.), Measurement Errors in Surveys, New York: Wiley.

Fowler, F.J., and Mangione, T.W. (1990). Standardized Survey Interviewing: Minimizing Interviewer-Related Error. Newbury Park, CA: Sage Publications.

Fracasso, M.P. (1989). Categorization of Responses to the Open-ended Labor Force Questions in the Current Population Survey (CPS). Proceedings of the Section on Survey Research Methods, American Statistical Association, 481-485.

Futterman, M. (1988). CATI Instrument Logical Structures: An Analysis With Applications. Journal of Official Statistics, 4, 333-348.

Gaertner, G.H., Cantor, D., Gay, N.L., and Shank, S.E. (1989). Tests of Alternative Questions for Measuring Industry and Occupation in the CPS. Proceedings of the Section on Survey Research Methods, American Statistical Association, 490-495.

Groves, R.M. (1989). Survey Errors and Survey Costs. New York: Wiley.

Groves, R.M., Berry, M., and Mathiowetz, N. (1980). Some Impacts of Computer Assisted Telephone Interviewing on Survey Methods. Proceedings of the Section on Survey Research Methods, American Statistical Association, 519-524.

Groves, R.M. and Mathiowetz, N.A. (1984) Computer Assisted Telephone Interviewing: Effects on Interviewers and Respondents. Public Opinion Quarterly, 48, 356-369.

Groves, R.M. and Nicholls, W.L. (1986). The Status of Computer-Assisted Telephone Interviewing Part H: Data Quality Issues. Journal of Official Statistics, 2, 117-134.

Hansen, M., Hurwitz, E., and Bershad, M. (I 96 1). Measurement Errors in Censuses and Surveys. Bulletin of the International Statistical Institute, 38, 359-374.

House, C.C. and Nicholls, W.L. (1988). Questionnaire Design for CATI: Design Objectives and Methods. In R.M. Groves, et al. (eds.), Telephone Survey Methodology, New York: Wiley.

Jabine, T., Straf, M., Tanur, J., and Tourangeau, R. (1984). Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines. Washington, DC: National Academy Press.

Kostanich, D.L. and Cahoon, L.S. (1994). Effect of Design Differences Between the Parallel Survey and the New CPS. CPS Bridge Team Technical Report 3, Washington, DC: Bureau of Labor Statistics.

Martin, E. (1986). Report on the Development of Alternative Screening Procedures for the National Crime Survey. Unpublished Report Dated January 15.

Martin, E. (1987). Some Conceptual Problems in the Current Population Survey. Proceedings of the Section on Survey Research Methods, American Statistical Association, 420-424.

Martin, E., Campanelli, P., and Fay, R. (I 99 1). An Application of Rasch Analysis to Questionnaire Design: Using Vignettes to Study the Meaning of ‘Work' in the Current Population Survey. The Statistician, 40, 265-276.

Miller, S. (I 994). What Would the Unemployment Rate Have Been Had the Redesigned Current Population Survey Been in Place From September 1992 to December 1993?: A Measurement Error Analysis. CPS Bridge Team Technical Report 1, Washington, DC: Bureau of Labor Statistics.

Mullin, P., Conrad, F., Sander, J., and Herrmann, D. (I 993). Cognitive Process Models of the Survey Interview. Unpublished Manuscript.

Nicholls, W.L. and House, C.C. (1987). Designing Questionnaires for Computer-Assisted Interviewing: A Focus on Program Correctness. Proceedings of the Third Annual Research Conference, Bureau of the Census, 95-101.

Palit, C. (I 987). Discussion of 'Questionnaire Design for Automated Interviewing'. Proceedings of the Third Annual Research Conference, Bureau of the Census, 128-129.

Palmisano, M. (1989). Respondent Understanding of Key Labor Force Concepts Used in the CPS. Paper presented at the Annual Meetings of the American Statistical Association, Washington, DC.

Payne, S. (1951). The Art of Asking Questions. Princeton, NJ: Princeton University Press.

Perreault, W.D. (1975). Controlling Order-Effect Bias. Public Opinion Quarterly, 39, 544-551.

Platek, R. (1985). Some Important Issues in Questionnaire Development. Journal of Official Statistics, 1, 119-136.

Polivka, A.E. (1994). Comparisons of Labor Force Estimates for the Parallel Survey and the Current CPS. CPS Overlap Analysis Team Technical Report 1. Washington, DC: Bureau of Labor Statistics.

Reason, J.T. (1988). Human Error. Cambridge, England: Cambridge University Press.

Rice, S.C., Wright, R.A., and Rowe, B. (1988). Development of Computer Assisted Personal Interview for the National Health Interview Survey. Proceedings of the Section on Survey Research Methods, American Statistical Association, 397-400.

Rothgeb, J.M. (1994). Revisions to the CPS Questionnaire: Effects on Data Quality. CPS Overlap Analysis Team Technical Report 2, Washington, DC: Bureau of Labor Statistics.
Shoemaker, H. H. (1993). Results from the Current Population Survey CATI Phase-In Project. Proceedings of the Section on Survey Research Methods, American Statistical Association.

ScfLwartz, N. (1987). Cognitive Aspects of Labor Surveys in a Multi-National Context. Paper Prepared for the OECD Working Party on Employment and Unemployment Statistics, June.

Sudman, S. and Bradburn, N. (1974). Asking Questions: A Practical Guide to Questionnaire Design. San Francisco, CA: Jossey-Bass.

Tanur, J. M. ed. (1991). Questions About Questions: Inquiries into the Cognitive Bases of Surveys. New York: Russell Sage Foundation.

T'homas, R. (1979). Discussion of 'The Limitations of Human Memory'. In L. Moss and H. Goldenstein, (eds.), The Recall Method in Social Surveys, University of London Institute of Education.

Thompson, J. (1994). Mode Effects Analysis of Labor Force Estimates. CPS Overlap Analysis Team Technical Report 3, Washington, DC: Bureau of Labor Statistics.

Thorslund, M. and Wdrneryd, B. (1985). Testing/Assessing Question Quality -- Some Swedish Experiences. Journal of Official Statistics, 1, 159-178.

Tiller, R. and Welch, M. (I 994). Predicting the National Unemployment Rate that the 'Old' CPS Would Have Produced. CPS Bridge Team Technical Report 2, Washington, DC: Bureau of Labor Statistics.

Tourangeau, R. (1984). Cognitive Sciences and Survey Methods. In T. Jabine, et al., Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines, Washington, DC: National Academy Press.

Turner, C. and Martin, E., eds. (1984). Surveying Subjective Phenomena. New York: Russell Sage.

U.S. Bureau of Labor Statistics (1988). Response Errors on Labor Force Questions Based on Consultations with Current Population Survey Interviewers in the United States. Paper Prepared for the OECD Working Party on Employment and Unemployment Statistics, June.

U.S. Department of Commerce, Bureau of the Census and U.S. Department of Labor, Bureau of Labor Statistics (1976). Concepts and Methods Used in Labor Force Statistics Derived From the Current Population Survey. Current Population Reports. Special Studies Series P-23, No. 62. Washington, DC: Government Printing Office.

van Bastelaer, A., Kerssemakers, F., and Sikkel, D. (1988). Data Collection With Handheld Computers: Contributions to Questionnaire Design. Joumal of Official Statistics, 4, 141-154.

c:\cps\copafs94.doc