CONDUCTING COGNITIVE INTERVIEWS TO TEST SELF-ADMINISTERED AND TELEPHONE SURVEYS: WHICH METHODS SHOULD WE USE?

Susan Schechter, National Center for Health Statistics, Johnny Blair, Janet Vande Hey, Univ of Maryland
Susan Schechter, NCHS, 6525 Belcrest Rd., Hyattsville, MD 20782

Key words: Pretesting, Mode, Cognitive Interviewing

Introduction

For more than 10 years now, the interdisciplinary efforts of survey methodologists and cognitive scientists have stimulated interest in establishing cognitive pretesting of questionnaires as a standard component of survey research. By applying cognitive psychology techniques to develop and test data collection instruments, survey researchers continue to improve and expand methods used to interview small numbers of subjects in a laboratory environment in order to identify questionnaire problems (Lessler, Tourangeau, and Salter, 1989; Willis, Royston, and Bercini, 1991; Jobe, Tourangeau, and Smith, 1993). These problems have typically been conceptualized in terms of a response model that begins with comprehension, then retrieval of information from memory, formulation of an answer and selection of a response category (Tourangeau, 1984; Willis et al., 1991). Cognitive interviews, like the survey questionnaires they test, can be administered in a variety of ways, and researchers must carefully consider pretest design decisions (Forsyth and Lessler, 1991; DeMaio, Mathiowetz, Rothgeb, Beach, and Durant, 1993; Beatty and Schechter, 1994). The focus of this paper is recent research on the different modes, methods, and interviewing techniques to use when conducting cognitive interviews to pretest self-administered and telephone surveys.

In the 1980's, household surveys conducted during a personal visit served as the framework for development of cognitive pretesting methods. During the same span of time, significant shifts were occurring in the expansion of mail and telephone modes of survey administration as well as the rapid expansion of new computer-assisted methods of data collection (Dillman, 1978; Groves, Biemer, Lyberg, Massey, Nichols, and Waksberg, 1988). Even though ground breaking changes in data collection methods were taking place, cognitive methods research remained focused mostly on face-to-face, paper and pencil mode. In instances where survey researchers wanted to study mode effects on measurement error, investigations were conducted in field experiments rather than in the cognitive laboratory.

Conducting cognitive pretest interviews without regard to the planned mode of survey administration also fit the cognitive interviewing model in psychology. In cognitive psychology laboratories, interviews with subjects were, and still are, primarily done face-to-face (Ericsson and Simon, 1993; van Someren, Barnard, and Sandberg, 1994), although there certainly are other laboratory research approaches, i.e., computer-interactive tasks, that would not include a face-to-face cognitive interview. Clearly, though, the think aloud interview used in cognitive psychology which was adapted for questionnaire pretesting is conducted in a personal interview (Jabine, Loftus, Straf, Tanur, and Tourangeau, 1984; Jobe and Mingay, 1991).

Therefore, it was no surprise that a review of the literature on current practices of cognitive pretest interview methods found that the predominant mode used in cognitive laboratory testing is face-to-face regardless of the planned mode of the survey (Willis et al., 1991; DeMaio et al., 1993). The literature does contain suggestions for how to utilize and vary cognitive laboratory methods depending on research objectives (see Forsyth and Lessler, 1991, for a taxonomy of different methods to use). Some work has also addressed the issue of mode in the laboratory environment (Bates and DeMaio, 1989; Gower and Dibbs, 1989; Jenkins and Dillman, 1994; Schechter, Beatty, and Block, 1994). However, guidelines for developing a cognitive laboratory protocol specific to a given mode of survey administration appeared to be absent in the literature. Filling this void was the impetus for our research.

The goals of this study were to: (1) identify different methods available when conducting cognitive testing of face-to-face, telephone, and self-administered surveys; (2) develop mode-specific guidelines for laboratory testing of telephone and self-administered surveys; and (3) determine which cognitive methods and interview techniques would be most useful when testing a telephone or self-administered survey. A three-phase study was designed. First, we conducted a survey of research organizations to determine whether laboratory methods differed by mode of survey administration. Second, we conducted focus groups with psychologists and survey methodologists involved in questionnaire design and the application of cognitive psychology methods to pretesting. Third, we planned a series of laboratory experiments designed to pilot test and evaluate new methods to pretest self-administered questionnaires and telephone questionnaires. Following is a brief description of the first two phases; the remainder of the paper is devoted to Phase III methods, results and implications thus far.

Phase I: How does the current practice differ by mode?

A convenience sample of organizations was drawn from four sources: (1) personal knowledge of organizations that conduct cognitive interviews; (2) organizations included in the Blair and Presser (1993) study; (3) organizations identified during a literature review; and (4) organizations referred to us by respondents during the study. Thirty-three organizations were selected for the sample; seven of these were not included in the final study either because of nonresponse or because cognitive laboratory pretesting methods were not being used. The final sample included 26 organizations: 5 government statistical agencies; 15 academic survey research organizations; 6 private sector firms. The respondent was the person who knew the most about cognitive interviewing in the organization. Respondents were first asked a series of questions about the pretesting of surveys to be administered in the face-to-face mode. They were then asked how cognitive pretesting methods differ when testing self-administered and telephone surveys.

When testing self-administered surveys, nearly all organizations reported that the cognitive interviews are done face-to-face. However, half of these organizations have laboratory subjects first complete the survey without interviewer interaction, followed by the interviewer conducting retrospective probing or a debriefing. The remaining organizations conduct the more typical face-to-face cognitive interview using techniques such as think aloud, read aloud, and probing while the subject completes the questionnaire.

When testing telephone surveys, more than half of the organizations said they conduct cognitive interviews by telephone, and the remainder conduct them face-to-face. Respondents who said their organization conducts cognitive interviews by telephone report using probing (both concurrent and retrospective) techniques as well as think alouds and debriefings.

No organization reported using a standard for selecting particular laboratory methods or interview techniques based on the mode of survey administration. Face-to-face was the preferred test mode, except when recruiting subjects was too difficult or conducting face-to-face interviews was too expensive. In those cases, respondents said their organizations conduct cognitive interviews by telephone, regardless of the planned mode of survey administration.

We concluded from Phase I that there are no established guidelines regarding test modes, methods, or techniques when designing cognitive research to pretest self-administered or telephone surveys,. In addition, the meaning and use of terms when describing cognitive research projects are not consistent or always shared among survey research organizations. For purposes of this project (and hereafter in this paper), we found it useful to define and standardize terms. Thus, *mode* is used to describe whether respondents answer survey questions in a face-to-face or telephone interview, or whether they answer the questions themselves with no interviewer present. Survey mode and test mode can each be face-to-face, telephone, or self. *Method* refers to the manner or means used to study cognitive processes when answering questions. Examples of methods are expert review, behavior coding of a field interview, a cognitive interview, a debriefing, and a focus group discussion. *Techniques* are those procedures used during a cognitive interview to study the response process. Common techniques are think aloud, probing (concurrent and retrospective), and vignettes.

Phase II: What did the experts say?

We conducted two focus groups with ten recognized experts in cognitive and/or survey research. The objectives were to discuss cognitive laboratory research from three perspectives: (1) the role of survey mode; (2) theoretical considerations for taking mode into account; and (3) the efficacy of particular laboratory methods for telephone or self-administered surveys.

One outcome of the focus groups was consensus that it does make sense to vary laboratory test mode according to the planned mode of survey administration. However, there was disagreement about the degree to which it should be varied,. Discussion centered on defining the purpose of the laboratory testing. If the purpose of the cognitive interview is to finalize survey questions prior to fielding, then simulating mode (and other field conditions for that matter) may be very important. But, if the purpose is to study cognitive difficulties that may cause response error, then simulating mode of cognitive interview with mode of survey administration may be less important, and in some cases, detrimental to accomplishing research objectives. Most participants agreed that to study comprehension of questions or words in questions, face-to-face test mode is best regardless of planned mode of survey administration.

Participants noted that the laboratory environment itself causes a difference in the respondent's survey task, which may undermine the need to account for survey mode. In contrast to survey respondents, lab subjects generally have more time to answer a question, have fewer distractions than at home, are more motivated due to voluntary participation and incentives, and usually engage in conversation during the cognitive interview. Consequently, interpretation of lab findings, particularly as they relate to mode concerns, may be difficult and error-prone. For example, if an interviewer observes a subject completing a self-administered questionnaire, how should the observed behavior be interpreted and evaluated? The subject may look confused, flip pages back and forth, and erase or make other corrections. Yes, the interviewer can document the observations but the basis for the observed behaviors is often unclear.

Both focus groups had concerns about interfering with the task of completing a self-administered questionnaire. Asking a subject to read and think aloud while answering questions takes the demands of a survey task and increases the cognitive load. As one participant stated, "If you do a face-to-face interview of a self-administered questionnaire, the nature of the task, reading and answering yourself with no interviewer interaction, is changed too much." Another said, "If you're doing concurrent think aloud, that [technique] may have the effect of really being disorienting to subjects. The person may miss skip patterns that they wouldn't necessarily miss if you'd left them alone to go through the whole questionnaire first."

A list of potential laboratory projects to investigate the testing of self-administered and telephone questionnaires was generated by the two groups. When testing a self-administered questionnaire, the presence or absence of the interviewer when a subject completes a questionnaire was thought to be the critical research design factor. When testing telephone questionnaires, conducting at least some cognitive interviews by telephone was strongly supported.

Phase III: Pilot study to test a self-administered questionnaire in the laboratory

Several laboratory experiments were conducted to investigate different methods for testing self-administered and telephone questionnaires. In this paper, only one of the experiments is discussed. The pilot study reported here sought to determine whether cognitive methods used to study the response process could be extended to understanding the cognitive demands and tasks unique to self-administered questionnaires.

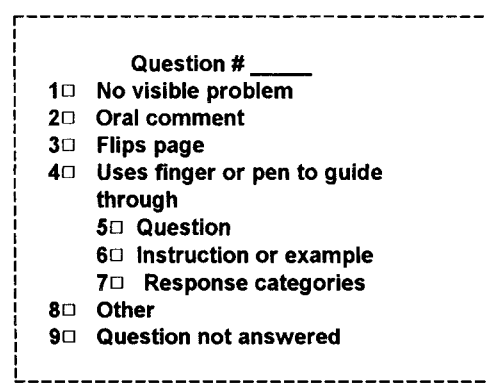
Methods

Sample: Forty-five subjects were recruited through volunteer responses to newspaper advertisements, flyers, telephone screening, and word-of-mouth. In Spring 1996, 30 interviews were conducted at the NCHS Questionnaire Design Research Laboratory and 15 were conducted at the University of Maryland Survey Research Center. Four experienced cognitive interviewers and two new cognitive interviewers were trained to administer the research protocol. All interviews were conducted face-to-face. Fifty-seven percent of subjects were female and 43% were male. Ages ranged from 19 to 68 with half of the subjects in the 35 - 54 age range. Eighty percent of the subjects had some college or more; the remaining 20% had completed the 12th grade. Three-fourths of the subjects reported that they were employed for pay the week prior to the interview and the remaining were students, retired, or homemakers.

Instrument: A self-contained section of a mail questionnaire that had already been fielded was used. Reasons for selecting the particular section of the questionnaire were the frequent use of respondent instructions, the requirement for subjects to refer to an appendix to answer one question, and a potentially unclear format for entry of numerical responses. Most of the questions were about work status such as whether the subjects worked full time or part time, the kind of work done, and so on. The instrument pilot tested had an introductory cover page of general instructions, 17 questions, and an appendix containing a three-page listing of job codes. Of the 17 questions, 12 contained an instruction (e.g., mark yes or no for each), 5 contained a skip pattern, and 6 contained definitions and/or examples.

Probing: The pilot study examined whether probing was effective when testing self-administered questions. Therefore, a split ballot experiment was conducted and subjects were divided into a "probe" group (n=23) and a "no probe" (n=22) group. The interviewer's verbal instructions to all subjects were (1) to complete the questionnaire as if they had received it in the mail at home; (2) to save any questions they had until they were finished; and 3) to expect that the interviewer might take some notes. Interviewers told the probe subjects that they might be asked a few questions during the interview, and that if that happens, to try to answer without looking back to a particular question. Interviewers also administered a short set of debriefing questions to all subjects after completion of the form.

Behavior coding: The study also examined whether subject behavior could be coded by a cognitive interviewer during the interview. The subject's copy of the questionnaire had the survey questions on the left half of the page with the right half of each page blank. The interviewer's copy contained a behavior coding box (see below) on the right half of the page, directly across each question. Interviewers were trained to observe the subject as he/she went through the questionnaire, and to check at least one code per question. Behavior coding was completed for both the "probe" and "no probe" groups.



Results[1]

Probing: Three types of probes were administered to the probe group. The first type asked for the recall of a questionnaire instruction. For example, the cover of the questionnaire contained five instruction points that addressed completion of the form. After the subject read the instructions and turned the page to start the questionnaire, interviewers probed with, "In your own words, can you tell me what the instructions said?" No subject recalled each of the five instruction points. Recall of a given instruction ranged from a high of 48% of subjects who recalled "when answering questions, please use an X" to a low of 35% of subjects who recalled "use a pen or pencil."

The second type of probe asked for recall of a definition provided as part of a particular question. Responses to one of these probes ("Do you recall how part-time work was defined?") revealed that 40% of the probe subjects recalled the definition accurately, while 27% said they did not recall the definition and 20% said they did not read it (because it didn't apply to them). In a similar example, interviewers probed with "Do you recall what the question meant by principal employer?" In this case, 61% of the subjects recalled the definition accurately, 33% recalled the wrong definition, and 6% reported not reading the definition at all.

The third type of probe sought to identify format problems. For example, one question required the subjects to match their current job to one of more than 200 codes contained in a listing. After answering this question, interviewers probed with "Would you say that the job codes list was pretty easy to use or not too easy?" All but three probed subjects said the list was easy to use. An analysis of completed questionnaires later found that a third of the subjects selected either an inaccurate code or selected the general catch-all code rather than a more specific code found to be available

Behavior coding: Coding of observed behavior revealed 71% of all subjects used their finger or pen to guide through either a question, an instruction, or a response category. This behavior was most frequent (28% of all subjects) for a question that contained a lengthy response listing for which subjects were to mark Yes or No for each. Not including responses to probes, 47% of all subjects were reported to make at least one comment during the interview. Close to half of the comments were in the form of a question to the interviewer ("Can I ask you a question?") and the other half were mostly reading the question aloud to themselves, referring to confusion ("I must have made a mistake.") or confirming understanding ("Oh, I see."). Among the remaining behavior codes, 33 % of subjects flipped a page back or forth when answering at least one question. The "Other-specify" box was marked for 69% of subjects. Analysis of the written interviewer entries revealed that more than half were notes indicating an observed skip pattern error, and the remaining were either reports of subjects putting their answer in the wrong place on the form, or reports regarding changes in subject demeanor (e.g., hesitation).

Subject debriefings: Eighty percent of subjects said the questionnaire was easy to fill out. Nearly 40% of all subjects said there were times that they were unsure what question was suppose to be answered. Subjects were shown four formats for entering numbers: one format was on the questionnaire and three formats were new and had not been seen before. Subjects were asked to indicate their preference(s). The format on the questionnaire was preferred by only 16% of subjects.

Interviewer self-debriefing: Interviewers filled out a short self-debriefing questionnaire immediately following the interview. Interviewers reports that the completion of behavior codes immediately following the interview was demanding and at times confusing, especially when combined with scripted probe administration. For 61% of the probe subjects, interviewers reported that probes were effective in identifying potential problems with questionnaire instructions. However, interviewers also reported that probing seemed to cause distraction for three-fourths of the subjects.

Identification of errors: Three sources of error counts were available in the pilot study: errors reported by interviewers in the "Other-specify" behavior code box; errors reported by interviewers in answering the self-debriefing questionnaire; and errors detected through an independent review of the subject's completed questionnaire. In the self-debriefing questionnaires, interviewers reported that 82% of all subjects made at least one error. The independent review found that 76% of all questionnaires contained an error. Entries on the interviewer's behavior coding box yielded the lowest count of errors (44%).

The independent review was the only method in which errors could be quantified and categorized. Among the 76% of subjects who made at least one error, 47% erred in following an instruction, 47% erred in following a skip pattern, 35% erred in formatting an answer, 24% provided

---

Data results from the subject questionnaires are not presented here as the paper's focus is on the testing methods used. Also, due to the small number of total cases, no statistical tests for significance were done on differences between the probe and no probe groups. We view these results as qualitative data to suggest whether there are any differences between the groups and in what direction the differences might be.

an obviously incorrect answer, and 15% failed to answer a question they should have. Note that subjects could make more than one error.

Discussion

Ways to test the cognitive burdens unique to self-administered questionnaires (e.g., instructions, format, and so on) were the focus of this pilot study. Methods to behavior code field interviews were adapted to a face-to-face cognitive interview to permit the interviewer to code subject behaviors during completion of the questionnaire. Half the subjects were asked cognitive probes during the completion of the questionnaire and half were not probed, in order to examine the usefulness of this interview technique. The probes and follow-on debriefing questions were targeted toward identifying cognitive difficulties subjects may have had in understanding and following instructions and format. Each completed questionnaire was independently reviewed and analyzed for errors to determine conceptually whether a review of fielded questionnaires would be of value as a method used in cognitive pretests.

The review of completed questionnaires allowed for a detailed categorization of error-type. However, it provided only minimal understanding of the information processing that led to the error. While obvious errors may well indicate instrument flaws, correctly completed questions and apparent correct following of instructions does not indicate the absence of problems. It may well be that some respondents reach the correct answer or follow the intended instructions despite the questionnaire design. In non-laboratory conditions, where respondents may be less motivated to work hard to do the tasks correctly, more errors may emerge.

We did not expect that 76% of reviewed questionnaires would contain an error. But we did predict and find that the no probe group tended to make more errors than the probe group. We suspect that the lower number of errors for the probe group is related to the probing, which we thought would cause subjects to pay more attention as they went through the questionnaire. Interestingly, the self-debriefing of the interviewers revealed that the interviewers perceived the probing as causing distraction, which one could reasonably argue would lead to more errors.

Behavior coding during a cognitive interview: Behavior coding provided a means for the interviewer to easily report behavior and highlight a potential problem to refer back to in the debriefing. It also provided a means to quantify what is happening during the process of a subject filling out a questionnaire. Codes were not intended to be error indicators. Rather, codes that were thought to reveal clues regarding the information processing tasks involved in completing a self-administered questionnaire were selected.

A drawback in behavior coding as used in this study was that most subjects were either vaguely aware that the interviewer was doing something, or acutely aware that the interviewer was making notations specific to the completion of a question. The interviewers were trained to put their copy of the questionnaire on a clipboard, and sit in a position to prevent subjects seeing what notes the interviewer was taking. It is unclear to what extent the subjects were aware of the interviewers' task. In retrospect, this should have been asked as part of the subject debriefing. Having the interviewer behavior code from behind a one-way mirror might work to measure some behaviors, but it would be difficult to focus on the subject and the questionnaire at the same time.

Along these lines, it would be interesting to see if more relevant information about non-verbal behavior could be captured by a observer rather than the interviewer. This observer could either be in the interview room, or could watch a videotape. This pilot study sought to determine whether interviewers could in fact code behaviors while conducting a cognitive interview. Results indicate that while an interviewer can do the coding, some important non-verbal behaviors may be missed due to the complexity of competing interviewer tasks.

Last, the behavior codes themselves were not as informative as anticipated. A different listing of codes would likely produce more meaningful data. For example, rather than use a general "oral comment" category, a more specific breakdown such as "oral comment indicating confusion" and "oral comment acknowledging error" could be offered.

Effectiveness of scripted probes: For subjects in the probe group, probes were scripted and strategically placed to detect specific problems with understanding and interpreting instructions and format. Probes were generally useful and provided valuable information on questionnaire instructions that were confusing or unclear. Not allowing any spontaneous (unscripted) probes for either group, and using only scripted probes for one of the groups clearly limited interviewer flexibility in following up of expressed problems or asking about an unanticipated subject behavior. However, because probing has such an impact on changing the nature of the task (and could potentially create more cognitive difficulty), frequency of probing was restricted and controlled. Allowing the interviewers to probe further would perhaps have revealed additional information about problems subjects were having.

A review of the probe group responses indicated that as subjects were probed, their responses to the probe questions became more accurate. We cannot conclude from that finding that the subjects actually were more accurate in completing their questionnaire, since the probes were not designed to measure question errors, but rather errors in understanding and interpreting instructions and definitions.

(In other words, just because the subject failed to accurately define part time work when probed does not mean that the subject answered the survey question incorrectly.) But this does suggest that subjects began to pay more attention to instructions and definitions in anticipation of further probes. Some probes to assess accuracy of responses would have been useful. More random probing (rather than the pattern used in this study which was to probe at the turn of the page) might be an effective approach worth considering.

Debriefing information from subjects and interviewers: The purpose of the debriefing questions was to determine how much additional information could be obtained about the instructions and format once the interview was finished. From a practical standpoint, being able to discuss with subjects what they thought about a given format and their view on different format options provided lots of clues about the visual processing that respondents go through in adapting to question completion tasks. An unexpected contribution of the interviewer's self-debriefing was the identification of initial errors that had been corrected later by the subject. These were not captured in either the independent review of the questionnaire or the behavior coding of a question because the question appeared to be completed correctly. Assessing when a respondent corrects a response may be a valuable measure and methods to do so should be pursued in further studies.

Stage of questionnaire development may guide mode, methods and techniques selected: The instrument used in this study was a final formatted questionnaire. We chose this instrument because it was representative of a "close to final" draft of a self-administered questionnaire. This is important in testing a self-administered questionnaire as some progress needs to first be made in instructions, formatting, layout, designing symbols, and so on. Testing an instrument in later stages of design is in contrast to what is typical in cognitive laboratory research. Many times the questionnaires tested in a face-to-face think aloud cognitive interview are in very rough draft format, perhaps not even ordered in a logical and systematic way. In addition, the appearance of the instrument is irrelevant as it is only the interviewer who is looking at the instrument. Intuitively, it seems that cognitive testing of self-administered questionnaires would be most useful when used in testing a questionnaire in an advanced stage of design. This allows the testing to focus on developing and evaluating how subjects understand and process the many instructions and visual symbols (such as arrows) found on self-administered questionnaires.

However, the pilot study also provided evidence that using cognitive laboratory methods to study cognition involved in processing questionnaire layout, format, and ease of use can be effective in designing self-administered instruments. In fact, the subject debriefing demonstrated that subjects can look at different sorts of answer formats, indicate their perceptions and preferences, and even give reasons for their preferences. The point here, though, is that we would not suggest "mixing" this sort of testing, with testing of questions themselves. Rather, the laboratory could serve as a testing site for early development of the best visual structure to an instrument, without regard to the types of questions. Work by Jenkins and Dillman (1994) supports this notion too.

Implications for future research: The pilot study results provided more information about questionnaire flaws and subject difficulties than about information processing during self administration. However, this is mainly a function of how the probes were scripted and the absence of think aloud protocols in this very preliminary research. It appears to be difficult to collect evidence about information processing, questionnaire flaws and potential respondent difficulties in the same interview. This paper has not presented the many observations the interviewers made regarding ways to improve the questions to address cognitive difficulties the subjects exhibited. In fact, it was a challenge for experienced cognitive interviewers to not focus on question problems, and rather, to focus on the characteristics of a self-administered questionnaire that were under study. It may be that using a mix of techniques within a given study -- think alouds and extensive probing in a portion of pretest interviews and a focus on observation and debriefings in the other portion -- would be a more reasonable approach to addressing questionnaire and format issues. Typical practice seems to choose one protocol for the entire pretest, but using two or more protocols may provide benefits in improving all aspects of a survey instrument. This is a empirical question worth testing in the future.

The errors observed in the independent review can, at a minimum, be indicative of problem points in the instrument. It may be that for self-administered questionnaires, the role of the laboratory may be maximized in two ways. First, interviewing subjects about instrument format, appearance, and layout, as well as ease of instructions, may be a valuable first step in the questionnaire design process. Second, improvements to the questionnaire may be identified by conducting interviews after a mini field pretest in which questionnaires are mailed to recruited subjects, sent back to the laboratory, and analyzed. Subjects would then be brought to the lab for further testing of questionnaire versions. Certainly, convincing evidence of problems based on reviewing completed instruments can come from conventional pretest methods. But the laboratory could be further utilized in the resolution of those problems and the testing of proposed solutions.

Finally, introduction of behavior coding by an interviewer while in the laboratory is useful and feasible. It

does add a burden to the interviewer, but in some ways, can expedite the process by which the interviewer documents perceived or obvious problems in answering questions. Clearly, consistent coding and interpretation of non-verbal behaviors are problematic, but it seems worth refining and retesting again. In addition to changing or expanding categories in future studies, another thought is to allow the interviewer to probe spontaneously about reasons for the observed behaviors. Last, if later work confirms that subject behaviors are affected by simply knowing the interviewer is coding a behavior after each answer, then the behavior codes could also be done randomly. Though some information would be lost, this may be worth a reduction of "interviewer-presence effects."

More generally, laboratory experimental research of this type requires considerably more cognitive interviews than would typically be used to test a questionnaire for a survey. At the same time, one is concerned that results may be (in part) artifacts of the particular questionnaire. For testing of hypotheses about cognitive laboratory procedures, ongoing tests on a range of studies is necessary. Ways need to be found to "piggy back" this sort of research onto other surveys or methodological studies. Similarly, this type of research requires much more uniformity and detailed reports of the testing procedures than has been typical to date. While this initial research did not provide the guidelines hoped for, it provided a rich source of information on ways to improve our testing methods in the future.

Acknowledgments: The authors gratefully acknowledge the valuable advice and support given by Dr. Stanley Presser. We also thank Elena Tracy, U of MD Survey Research Center, for the timely and accurate coding and processing laboratory data. Finally, we want to extend our deep appreciation to the NCHS and U of MD SRC staff that served as subject recruiters, cognitive interviewers and/or reviewers of this paper.

References

Bates, N. and DeMaio., T. (1989). "Using Cognitive Research Methods to Improve the Design of the Decennial Census Form." Proceedings of the Fifth Annual Research Conference, Washington DC: Bureau of the Census, 267-285.

Beatty, P. and Schechter, S. (1994). "An Examination of Mode Effects in Cognitive Laboratory Research." 1994 Proceedings of the Section on Survey Research Methods, American Statistical Association, 1275 - 1280.

Blair, J. and Presser, S. (1993). "Survey Procedures for Conducting Cognitive Interviews to Pretest Questionnaires: A Review of Theory and Practice." 1993 Proceedings of the Section on Survey Research Methods, American Statistical Association, 370-375.

DeMaio, T., Mathiowetz, N., Rothgeb, J., Beach, M., and Durant, S. (1993). "Protocols for Pretesting Demographic Surveys at the Census Bureau." Report of the Pretesting Committee, Washington DC: Bureau of the Census.

Dillman, D. (1978). Mail and Telephone Surveys: The Total Design Method. New York: Wiley.

Ericsson, K. and Simon, H. (1993). Protocol Analysis: Verbal Reports as Data. Cambridge, MA: MIT Press.

Forsyth, B. and Lessler, J. (1991). "Cognitive Laboratory Methods: A Taxonomy." In Biemer, P., Groves, R., Lyberg, L., Mathiowetz, N., and Sudman, S., eds. Measurement Error in Surveys. New York: Wiley.

Gower, A. and Dibbs, R. (1989). "Cognitive Research: Designing a 'Respondent Friendly' Questionnaire for the 1991 Census." Proceedings of the Fifth Annual Research Conference, Washington DC: Bureau of the Census, 257-266

Groves, R., Biemer, P., Lyberg, L., Massey, J., Nichols, W., and Waksberg, J., eds. (1988). Telephone Survey Methodology. New York: Wiley.

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

Jenkins, C. and Dillman, D. (1994). "The Language of Self-administered Questionnaires as Seen Through the Eyes of Respondents." Presented at the 1994 Council of Professional Associations on Federal Statistics, Bethesda, MD

Jobe, J. and Mingay, D. (1991). "Cognition and Survey Measurement: History and Overview." Applied Cognitive Psychology, 5, 175-192.

Jobe, J., Tourangeau, R., and Smith, A. (1993). "Contributions of Survey Research to the Understanding of Memory." Applied Cognitive Psychology, 7, 567-584.

Lessler, J., Tourangeau, R., and Salter, W. (1989). "Questionnaire Design Research in the Cognitive Research Laboratory." Vital and Health Statistics, 6(1), Hyattsville, MD: US Department of Health and Human Services.

Schechter, S., Beatty, P., and Block, A. (1994). "Cognitive Issues and Methodological Implications in the Development

and Testing of a Traffic Safety Questionnaire." 1994 Proceedings of the Section on Survey Research Methods, American Statistical Association, 1215 - 1219.

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

van Someren, M., Barnard, Y., and Sandberg, J. (1994). The Think Aloud Method. London: Academic Press.

Willis, G., Royston, P., and Bercini, D., (1991). "The Use of Verbal Report Methods in the Development and Testing of Survey Questionnaires." Applied Cognitive Psychology, 5, 251-267.