

DISCUSSION

Elizabeth A. Martin, U. S. Bureau of the Census
Washington, D. C. 20233

KEY WORDS: Questionnaire design, cognitive aspects of survey methodology, pretesting

I'd like to start by talking briefly about the purpose of the research activities reported in this session, and then discuss the individual papers.

These studies are part of a program of research being conducted collaboratively by the Bureau of Labor Statistics (BLS), the Census Bureau, and several research organizations under contract to BLS. The primary goal is practical: to redesign survey questionnaires and procedures. For this purpose, the research is useful insofar as it identifies problems in current questionnaires and procedures and suggests improvements for revision.

As I see it, we also have a broader aim, which is to develop a better way of designing and testing questionnaires. Presently there does not exist, and we are striving to create, a methodology by which we can understand the properties of survey instruments. We are trying to develop tools and methods that allow us to critically examine the assumptions implicit in survey questions and procedures. The aim is to demystify the response process, to open up the black box by taking a close look at survey questions and procedures from a respondent's point of view. We do this by applying ideas and methods from psychology and other social science disciplines to identify and analyze the underlying causes of errors in surveys. Of course, in developing and applying new methodology, we must take care not to ignore sound statistical and survey principles. For this broader purpose, the research is useful insofar as it contributes to methodological development and suggests new approaches to understand causes of errors in surveys.

With this as background, let's turn to the papers themselves. Four of the 5 papers in this session are concerned with the Current Population Survey (CPS), and one with the Consumer Expenditure Survey. I'll be discussing them primarily in terms of their practical utility in providing useful information for survey redesign, but also will consider them from a methodological point of view.

In her paper, Maria Fracasso evaluates the categorization of responses to open-ended questions in the CPS. She has rightly focused on an important problem, because a number of important classifications in CPS depend on respondents' answers to open-ended questions. For example, classification of a respondent as unemployed, and the determination that a person works part-time for economic reasons, are both based on responses to open-ended questions. Her approach is to examine a sample of verbatim

responses to open-ended questions and suggest revisions to the code categories. She tests the revisions by asking interviewers to code samples of responses using the revised coding schemes. In general, her suggestions for revision appear to be useful ones. However, there are several serious problems with her study. Her samples of verbatim responses are too small (only 50 for each question), and do not represent samples in any statistical sense. (They were taken from various sources: verbatim responses to the same questions asked in another survey; hypothetical answers made up by subjects in cognitive laboratory tests; and verbatim responses coded as "other" in CPS.) These ad hoc "samples" do not provide an adequate basis for evaluating the reliability or validity of the existing code categories or for recommending revisions to them. A second problem is that she evaluates her revised codes using the same verbatim responses she used to create them. The new codes are almost bound to show improvements, since they have been designed to capture whatever is idiosyncratic about those particular responses. It is standard survey practice to evaluate a set of code categories using an independent sample of responses which was not used to create the categories in the first place.

In his paper, Lawrence Boehm considers the reliability of proxy responses in CPS. Again, he has identified an important problem, since CPS collects labor force data about all household members from one respondent. He draws two conclusions from his results. First, he finds rather low levels of agreement between proxy and self-reports of labor force status, and concludes that CPS proxy data are unreliable. The problem is that his study cannot distinguish simple response unreliability from possible proxy effects. He cannot rule out the possibility that low levels of proxy-self agreement are due entirely to unreliability. His second finding is that proxy respondents' ratings of confidence in their ability to report reliably and their ratings of knowledge about the other person's job search were uncorrelated with accuracy. (Agreement with self-response is the criterion for accuracy.) That is, not only do proxy respondents not know much about labor force activities of other household members, but they don't know that they don't know. This conclusion is surprising, and a bit difficult to evaluate. He does not tell us exactly how he measured confidence and knowledge (confidence in what? knowledge about what?). It would be useful to look at the circumstances which produce agreement and disagreement (either self-self, or self-proxy agreement). Mr. Boehm's data show that there is more confusion between the categories "unemployed" and "not in the labor force" than between "employed" and either of the other two. This is an interesting result

and may provide a clue as to where the problems lie. He should also consult O'Muircheartaigh's (1986) analysis of CPS reinterview data, from which O'Muircheartaigh concludes that the reliability of the data (for both proxy and self-reports) varies according to a respondent's relationship to the household and depends on the clarity or robustness of the information a respondent is asked to report.

In their paper, Sherman Edwards, Roger Levine, and Sharon Cohany attempt to develop procedures for validating CPS respondents' reports of the hours they worked the previous week. There are several very promising ideas in this paper. One is their attempt to develop questions and procedures for self-validation, which they used to elicit a much more detailed and complete report of hours to verify respondents' initial reports. They combined this self-validation procedure with a partial record check, whereby they asked each respondent to report separately just those hours that his or her employer should have on record; this information was then compared to employer records. This mixed strategy of using various sources of information to evaluate respondent reports is a promising approach to the knotty problem of validation. On the other hand, their attempt to identify global causes of error (motivation, recall, comprehension, etc.) to explain discrepancies between different reports is less convincing. It seems unlikely that, in most cases, it will be possible to cleanly identify a single general cause of a particular survey error.

Gregory Gaertner, David Cantor, Nancy Gay, and Susan Shank direct their attention to a costly and error-prone operation in the Current Population Survey: industry and occupation coding. Previous research documents substantial unreliability in occupation and industry data, with high rates of coding error. In addition, the coding operation is costly and time-consuming. Questionnaire revisions which reduce the amount of coding thus potentially could improve quality and reduce costs at the same time, a very desirable outcome. Particularly in the case of industry, their testing shows some promise. They obtain more consistent industry data by slightly modifying the question currently used in CPS. However, they consider only a few alternative question wordings. In my view, they do not go far enough in developing different approaches based on careful examination of the sources of confusion and ambiguity in the codes. Their revised wordings do not appear to be based on in-depth knowledge of the substance of the industry and occupation classifications. Another limitation of the study is the fact that their interviews were not conducted by trained interviewers, so the results may not apply generally.

Finally, Leslie Miller and Theodore Downes-Le Guin consider the problems of recall and comprehension in the Consumer Expenditure Survey. These authors identify a number of problems in the wording and structure of the questions, and offer what appear to be useful and practical solutions, although they are untested. The authors might find it useful to review the research conducted to support the redesign of the National Crime Survey (NCS) (see Biderman, et al., 1986; Martin, et al., 1986). That research tested cuing as a strategy to stimulate recall, with positive results. The same approach may be useful in the Consumer Expenditure Survey, which, even more than NCS, poses a formidable recall task for respondents.

I had a minor quibble with this paper: the authors misinterpret part-set cuing, which does not imply relying "on some members of a category as cues to maximize the amount of information reported for all members of a category." Evidence on the part-set cuing effect implies just the opposite conclusion, that giving cues interferes with recall for noncued items in the same category.

As a group, these papers offer a number of useful suggestions for questionnaire revision. They also suggest promising new approaches to questionnaire design and testing. However, a couple of cautions are also in order. Most of these papers relied on very small numbers of respondents. The general approach of this type of research is to do more intensive study of a smaller number of respondents than is done in a typical survey. Nevertheless, there must be enough cases to support conclusions if statistical comparisons are made (as they were in these papers). Second, in attempting to find out what is wrong with surveys, it is important not to abandon what is right with surveys. In some instances, authors based conclusions on rather shaky assumptions and methods which were not consistent with good survey practice.

REFERENCES

- Biderman, A. D., D. Cantor, J. P. Lynch, and E. Martin. (1986). Final Report of the National Crime Survey Redesign Program. Washington, D.C.: Bureau of Social Science Research.
- Martin, E., R. Groves, J. Matlin, and C. Miller. (1986). Report on the Development of Alternative Screening Procedures for the National Crime Survey. Washington, D. C.: Bureau of Social Science Research.
- O'Muircheartaigh, C. (1986). "Correlates of Response Inconsistency in the Current Population Survey (CPS)." Proceedings of the Second Annual Research Conference. Washington, D.C.: U. S. Bureau of the Census.