

University of Wisconsin—Madison
MADISON, WISCONSIN 53706

Department of Sociology
Social Science Building
1180 Observatory Drive

To call writer:
Department (608) 262-2921
Office (608) 262-2868

March 1, 1987

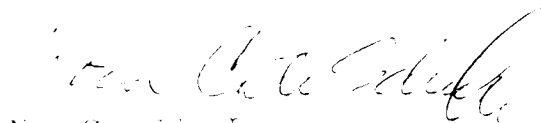
Cathryn S. Dippo
Senior Mathematical Statistician
Office of Research and Evaluation
Bureau of Labor Statistics
441 G Street N.W.
Washington, D.C. 20212

Dear Dr. Dippo:

I have enclosed my report from the Questionnaire Design Advisory Conference in January. As the introduction states, my suggestions focus on the Consumer Expenditure Interview Survey and the Current Population Survey. In some cases the comments are necessarily abbreviated; please let me know if I can expand or clarify any of these suggestions or if I can be of assistance in some other way.

I appreciated the opportunity to participate in the conference. I hope that the delay in the report does not cause any inconvenience.

Sincerely,



Nora Cate Schaeffer
Assistant Professor

Enclosures: as stated

Questionnaire Design Advisory Conference

Report — February 1987

Nora Cate Schaeffer

Department of Sociology

University of Wisconsin — Madison

This report summarizes my comments on each of the three surveys discussed at the Questionnaire Design Advisory Conference. The report has three principal sections, one each for the Consumer Expenditure (CE) Interview survey, the Consumer Expenditure (CE) Diary survey, and the Current Population Survey (CPS). Suggestions that are pertinent to more than one of the surveys are generally not repeated; the extensions across surveys are straightforward. Nevertheless, because many of the suggestions can be generalized to all three surveys, the sections are not completely independent. Most of the discussion concerns the CE Interview and CPS; the experiments on the CE Diary described at the conference incorporate changes I would otherwise have suggested.

My report focuses on a set of general issues and suggestions for research on those issues. In addition, I provide some specific comments on questions in the CPS Test Questionnaire. The report does not attempt to be comprehensive, nor does it comment on current forms of individual questions.

Specific research suggestions are identified by • in the text. Much of the research suggested can be accomplished, at least initially, using small geographically concentrated samples of 50-100 respondents. It is important, however, that these samples be heterogeneous; accurate reporting depends on social as well as cognitive factors, and social and cognitive processes may interact in their effects on accuracy. For example, which recall strategy is most effective may depend on the frequency of a behavior, and whether or not records are available may depend on how purchases are made (e.g., cash, check, or credit card). Clearly, social factors influence both the frequency of purchases and how they are

made, and the sampling frame for small-sample studies must encompass the kind of variation on important social variables that is expected in the full survey. Whether or not any of these small research projects would then be extended to a larger sample would depend on the outcome of the initial study.

Two needs are primary, although identifying priorities among the research issues outlined below is difficult since the projects vary in complexity and expense as well as in their potential impact on data quality. First, developing criteria to evaluate the success of improvements in question wording and questionnaire design is clearly important, and the problems that arise in this respect are different in each of the surveys. Second, developing both rules to select the best respondent with the best store of information and procedures to support that respondent's efforts are essential: the best recall procedures will fail with a respondent who never knew the answer or who does not believe that accuracy is desired.

CONSUMER EXPENDITURE INTERVIEW

Given the limits imposed by costs and by interviewer and respondent energy, the design of the CE Interview embodies a decision about whether to obtain information on a large number of expenditures knowing that the data may include guesses or hurried estimates or to collect information on a smaller number of expenditures attempting to obtain accurate information. The decision to sacrifice accuracy can be implicit in the design of a questionnaire: if the amount of information requested can be provided within the time available (or within the time that the interviewer or respondent expects the interview to last) only by relying on guesses or rough estimates, the interviewer and respondent may come to an understanding that precision is needed only where specifically requested (as it currently is for phone bills and utilities). Obtaining accurate information, on the other hand, requires an explicit decision: the goal must be systematically communicated to the respondent and the interviewer must provide guidance and training in achieving the goal. Some of the suggestions made below would lengthen the interview: they assume that the scope of the interview can be reduced in other ways to accommodate them. The monthly recall period and split forms, already included in planned experiments, may provide some of the reduction in scope needed in order implement other improvements.

The comments on the CE interview are presented in five sections: **Criteria for Evaluating Changes** discusses possible criteria for evaluating outcomes of experiments on the survey. **Interviewer-Respondent Interaction** consists of two subsections that discuss the issues of respondent motivation and respondent training and rapport. **Developing Flexible Data Collection Strategies** discusses respondent rules, use of records, and a self-administered form. **Improving Recall** focuses on recall strategies, the specific question forms used in the CE, and recall aids. The final section briefly discusses **Conditioning Effects**.

Criteria for Evaluating Changes

While the success of any proposed procedure should be evaluated in an experiment, the choice of an outcome measure is problematic. This section suggests several criteria: they vary in how expensive they would be to use and in how strong are the assumptions they require. Those that are less clearly indicators of improved accuracy should themselves be evaluated. Although time consuming, such an evaluation would make it possible to identify criteria that can be used in regular monitoring of reporting quality or added to ongoing relatively easily for special assessments.

Possible indicators of improved reporting are numbered and described below.

1. frequency of reporting or dollar value of purchases.

Relying on increased reporting assumes that most errors are omissions or underestimates. Increased reporting appears to be the criterion currently used most often in evaluating accuracy.

2. recall accuracy.

The accurate recall of items listed for the third recall month at the previous interview (referred to as the "item list" below) can be used to evaluate both telescoping and omission errors for a three-month recall period under the assumption that the original report is accurate. Of the six criteria mentioned here, this is the only one that offers a direct indication of the accuracy of reports.

- Using the transcribed item lists to evaluate relative accuracy, whether of an experimental treatment or of some alternative criterion, requires systematically coding all items reported in the current month and their categories (e.g., apparel, home furnishing, etc.). The subsequent interview would record whether or not the item was correctly mentioned as being purchased in the third recall month. The outcomes would include whether or not the item was omitted

or telescoped forward. After all prompting was completed, respondents would be asked when omitted items had been purchased; this permits distinguishing items telescoped backward from those forgotten, although there is likely to be considerable error in these reports. Finally, any items reported in the current interview but not reported in the earlier interview would be recorded; such information would be useful in examining the performance of different kinds of respondents for different items. (See the subsection on **Respondent rules** below.)

If incorporated into experiments planned for 1988-1992, studies that use this criterion can compare a one-month and a three-month recall. Studies using this criterion are likely to be expensive. Initially, studies using the item lists can (1) identify problems in using increases in average expenditure as a criterion and (2) evaluate the other criteria described below. If some aspects of the procedures developed to use these item lists in research can be routinized, using the lists for future research projects becomes simpler and cheaper.

Assuming that item lists will continue to be used in the future even if a one-month recall period is adopted after experiments are completed, the way in which these lists are transcribed and coded can be standardized so that they can be adapted to research use when needed.

3. use of records.

Records can be consulted for two reasons: to determine whether or not a purchase occurred and to determine how much the item cost. Using this criterion makes the assumption that accuracy increases when records are consulted.

Currently whether or not records are used for specific items is recorded only for utilities and telephone bills.

- Detailed analysis of this criterion would require recording additional informa-

tion: the classes of items for which records were consulted, what records were consulted (bills or checkbooks), and how often. The current control card item that asks about record use could be replaced by one that asked the rough proportion of items for which the respondent consulted records.

4. consulting others.

Similarly, this criterion assumes that accuracy increases when others are consulted, whether about the occurrence of purchase or about the price of an item.

- Detailed analysis of this criterion would also require recording additional information: the classes of items for which others were consulted, who was consulted, how often, and whether or not they used records.

5. response time or time to complete sections of the questionnaire.

Using this criterion assumes that accurate reporting—whether improved accuracy is due to reflection or remembering, to record checks, or to consultations with others in the consumer unit—takes time.

- Detailed analysis of this criterion would require recording the time spent on relatively small individual sections of the questionnaire: the length of time could then be standardized by the number of items answered. This criterion might be appropriate, for example, when evaluating the effects of techniques designed to increased respondent motivation and commitment.

6. respondent is householder or spouse.

This criterion assumes that compared to others, respondents in these roles will have more complete knowledge of the spending of more household members for more types of purchases. A variation is to ask respondents to identify the most knowledgeable respondent. (Otherwise knowledgeable respondents, however, may be biased reporters about certain areas such as the clothing or entertainment purchases of adolescents.)

Each of these criteria may be an appropriate indicator of reporting accuracy in different contexts. With the exception of the first two, the criteria merely specify interview conditions that may themselves tend to increase accuracy. The item lists can be used to evaluate any of the procedural changes suggested below. In addition, these lists can be used to evaluate the other criteria. For example, while using records and consulting other household members probably can be presumed to improve accuracy, the presumption is less strong for increased frequency or dollar value of reports, for response time, or for using the householder or their spouse as the respondent. (See the subsection on **Respondent rules** below.) If these other, less expensive, indicators are indeed associated with improved performance, they can be collected on an ongoing basis or periodically for specific studies.

Interviewer-Respondent Interaction

The suggestions in this section focus on the interaction between the interviewer and the respondent. Because development and testing of these procedures requires extensive feedback from respondents, the first stages of work can be done with relatively small samples. Evaluating the effectiveness of any resulting procedures, of course, requires a larger experiment.

Respondent motivation

Since respondents assume their role without much experience, it is usually necessary to explain what is wanted from them. This description should be simple and direct. At a minimum, information about the uses of the data, the importance of accuracy, and confidentiality needs to be presented systematically to all respondents. In addition to motivating the respondent, such a presentation can be designed to build a task-oriented rapport and to establish accurate reporting as the general goal of respondent training. Simply giving the respondent a brochure to read in their own time before or after the

interview may increase the interviewer's credibility but is unlikely to motivate the respondent. A brief (about five minutes) narrative emphasizing the uses of the data and the need for accuracy can be written around brochures such as *Your Responses are Vital* or *How the Census Bureau Keeps Your Information Strictly Confidential*. Alternatively, more visually attractive materials can be specifically designed for such a presentation. In either case, the presentation can be used to demonstrate that the data are perceived as important and to provide an opportunity to elicit the respondent's commitment to accuracy.

- The presentation script and materials should be systematically pretested to discover whether or not they succeed in (1) communicating the importance of the survey and its goals, (2) addressing respondents' concerns, and (3) holding the interest of respondents. The effectiveness of the presentation and commitment can be evaluated using one of the criteria described earlier.

The demographic and attitudinal correlates of nonresponse are necessarily difficult to study (see for example, DeMaio 1980; Stinchcombe, Jones, and Sheatsley 1981). In addition, however, little is known about why respondents who cooperate do so or about which motives are associated with increased accuracy. Studying respondent motivation is complicated by the fact that while respondents may know why they refuse, many probably do not know why they cooperate—except that they cannot think of a good reason to refuse. Furthermore, conventionalities and social desirability pressures may influence answers to questions about cooperation.

- For initial studies of respondent motivation, a small random subsample (about 100 cases) of CE respondents at a small number of locations could be selected for a brief in-depth interview about the respondent's participation in the survey after the last interview. The goal of such interviews would be to describe cooperative motives and obstacles to expressing these motives. Concerns of respondents revealed in such interviews could then be addressed systematically in the description of the survey given to new respondents. Even in such a small-scale study, an attempt should be

made to compare the motivations and personal characteristics of those who do and who do not provide accurate information.

Respondent training and rapport

The opening presentation begins respondent training. It should be followed by a description of the task, a statement of the importance of accuracy, suggestions about the strategies most likely to increase accuracy, for example, the suggestion that the task will be easier if the respondent collects records before the interview begins.¹ (See **Self-administered form** below.)

The interviewer training and the interview schedule should incorporate interviewer behaviors that reinforce the importance of accurate reporting and aid the respondent in meeting this objective.² Reinforcing accurate reporting involves giving positive feedback for desired behaviors, requesting that other answers be reconsidered, and encouraging the respondent to take the time necessary to recall the information requested. Reinforcing accurate reporting can be done most easily in an interview that covers less material than the current CE interview for two reasons. First, the feedback and recall time themselves lengthen the interview. Second, the task should be designed so that respondents can

According to comments made at the conference, current practice is for the interviewer to decide whether or not to ask the respondent to collect records before the interview. This lack of standardization is likely to increase interviewer variance. The interview schedule specifies that records be used for utilities and telephones, but the record requests are not made at the same time. In my own interview, no preliminary suggestion about using records was made and I assumed that this was deliberate, that the interviewer preferred that I not use records unless specifically requested. A global request for records at the beginning of the interview would have communicated the goal of the task more correctly, made responding easier, and been less alienating than the the current piecemeal approach which resulted in my making four trips upstairs to my files.

¹Some of the suggestions in this section draw on the work summarized in Cannell, Miller, and Oksenberg (1981).

Be given
a request
for records

complete it within a reasonable length of time; if respondents are sent a double message that both accuracy and speed are desired, they may heed the suggestion of speed. Ways of aiding the respondent in accurate reporting can be designed once more is known about (1) how respondents keep records and (2) the way in which lifestyle variables influence reporting accuracy. (See the subsection on **Use of records** below.)

Interviewers should be taught to help respondents and how to do it. Time should be allowed at the end of the initial bounding interview for the interviewer to explain the content of future interviews, to make suggestions about how to prepare for those interviews, and to discuss problems the respondent anticipates. For respondents who anticipate problems, one or more followup contacts by telephone between the bounding interview and the next interview may be necessary. In this call the interviewer asks what problems the respondent has been encountering in collecting receipts from household members or in otherwise keeping track of expenditures; the interviewer then discusses with the respondent ways of resolving these problems. In addition to its practical focus, this phone call from the interviewer may help maintain respondent motivation and reinforce the respondent's perception that accuracy is important.

In training respondents the interviewer actively contributes to a rapport in which the interviewer is perceived as someone who can help the respondent be an accurate reporter. Even when this rapport is highly scripted, however, interviewers may inadvertently undermine rapport.

- Observational studies of interviewers can identify nonverbal interviewer behaviors that communicate that speed is more important than accuracy. These behaviors may range from inconsistency in requesting record checks to being poised to record an answer before the respondent has formulated an answer. It is possible, for example, that effective recall requires temporary disengagement from the social demands of the interview; if the interviewer communicates that speed is desirable, respondents may quicken the interview by omitting these periods of disengagement. Such a study

of interviewers — perhaps fifty at different levels of experience — would involve bringing interviewers to a central location to conduct an interview that would be observed through a two-way mirror. After the interview, respondents would complete a self-administered form about their perceptions of the interview and the interviewer. Specific hypotheses would focus on the relationship between interviewer behavior and respondent perceptions and on that between respondent perceptions and an indicator of accuracy or quality.

Developing Flexible Data Collection Strategies

Because of the great variety in the kinds of records respondents keep and in the level of knowledge of respondents, data collection strategies that take this variability into account may both increase accuracy and decrease burden. Research is needed to develop respondent rules to identify the respondent who both has information needed and is in a position to cooperate. In addition, an understanding of how people keep and use records and of which items records are typically available for may suggest improved methods of asking about purchases. Finally, for some respondents, filling out a self-administered form or worksheet before the interview may make it easier to give accurate information.

Respondent rules

Information about household members' expenditures and having the time to be an accurate respondent are not uniformly distributed within households. Two kinds of research on respondents are needed. The first would use one of the criteria described above to compare how respondents in different household roles and in households with different compositions (e.g., householder, spouse of householder) differ in their use of records and their accuracy. Being a cooperative and accurate respondent may be primarily a function of lifestyle factors: an adult who keeps the household accounts or one who makes most purchases may be the most knowledgeable and accurate respondent. This research would

also examine the stability of respondents over the life of the panel: are the most accurate respondents likely to remain the respondent? Information about typical distributions of information and about how much variability there is in these distributions would contribute to (1) developing rules for selecting the best available household respondent and (2) developing suggestions for helping that respondent get information from other household members.

- In-depth interviews with all adult members of households in a small (100) heterogeneous sample of two-or-more person households would have the objectives of (1) identifying who in the household knows the most about what purchases, (2) determining whether or not there is consensus about screening items such as "who knows the most about all the purchases in this household," (3) identifying the kinds of purchases most likely to be known only to the purchaser, and (4) developing methods for systematically getting information from household members other than the respondent. These interviews could also discuss the use of records as described in the next subsection.

Use of records

Again, two types of research subjects are indicated. First, the extent to which accuracy is improved by record checks for items of different types can be evaluated. Second, information is needed on the ways in which households store information about different kinds of expenditures and the distribution of and variation in these methods. For example, in some households, most purchases are made by credit card; if receipts are not kept systematically, however, records about purchases may not be available until one or two months after the purchase. In other households, almost all purchases are made by check so that reviewing check registers may provide highly accurate information. A simple screening question about the use of check registers may decrease burden for respondents who use checks regularly and permit interviewers to spend more time with respondents

who must rely more heavily on recall.

- In-depth interviews with all adult household members in a small random sample of households about their accounting and record-keeping procedures could suggest alternative ways of collecting accurate information from respondents: interviewers could be trained to select the method that promised the most accurate information with the least burden to respondents.

Self-administered form (worksheet)

Some respondents, especially those with very busy schedules, may find participating in the survey easier if they can prepare for the interview in advance. The in-depth interviews suggested above can provide guidance for developing a worksheet that respondents can use to assemble and record some or all of the detailed information needed for the interview. The worksheet would resemble a self-administered questionnaire. If the worksheet were complete at the time scheduled for the interview, the interviewer would review the worksheet for completeness and administer a brief supplementary recall section as well as sections of the interview not covered by the worksheet. Respondents who had not completed the worksheet would be interviewed using a standard schedule. Such a worksheet would not be suitable for all respondents, but some would probably prefer it to a lengthy interview; for some, this data collection strategy will increase accuracy. If sent to respondents one week before the interview, the worksheet would provide the respondent the opportunity to assemble records in a way convenient for them and to review purchases with other household members who might not be present during the interview.

- Evaluation would focus on identifying which sections could be best adapted to a worksheet format, developing a procedure to identify which respondents might benefit from using such a worksheet, and comparing the accuracy of results obtained using the worksheet to that obtained in a conventional interview.

Improving Recall

Recall within even a one month time frame can be taxing; a quarterly recall period is more than three times as difficult. It is likely that in the experimental comparisons of recall periods currently planned, one-month recall will prove superior since confining recall requests to one month has the desirable effects of simplifying the recall task and shortening the interview. In any case, research on improving recall can supplement this experiment.

Recall strategies

The effectiveness of recall strategies probably depends both on the time frame and on the type of information being recalled. Recall strategies used for the kinds of information requested in the CE interview can be studied using observation and in-depth interviews that use techniques such as concurrent think-aloud interviewing. This research would identify the range of recall strategies used, propose recall aids and flexible procedures for using them, and develop hypotheses about which strategies increased accuracy. These hypotheses could then be pursued in larger experiments that examined whether or not the techniques improved accuracy.

- Again, interviews with a small heterogeneous sample drawn from the same frame as the CE interview sample is indicated: these respondents would not have to be CE respondents and could be restricted geographically for convenience. Separate studies would focus on different kinds of information — purchases, credit, and so on — to keep the size of the task manageable. Some interviews would be conducted in a central location where respondents could be observed through a two-way mirror; other interviews would be conducted in respondents' homes. The task would be unstructured for respondents in the observed condition: materials (such as scratch paper, a calendar, and a calculator) would be available in the room, but respondents would determine for themselves how to go about filling out lists of all purchases in the months specified. In both conditions, after respondents reported purchases, they

would be asked to describe how they went about remembering, what estimating and averaging strategies were used for which items, what would have helped them in remembering, and how accurate they judge their reports to be.

By describing the variation among respondents in recall strategies, this research would provide a basis for developing flexible procedures that interviewers could be trained to use in helping respondents. For example, respondents probably vary both in how helpful they find a calendar and in how they use one if it is available: some respondents may find public events useful anchors while others prefer personal landmarks. Furthermore, effective use of a calendar probably requires a rapport that permits the respondent to disengage from the social aspects of the interview: a respondent who feels pressure from the interviewer to respond quickly may actually feel increased performance anxiety if offered a calendar.

Specific question forms

A better understanding is needed of the recall and estimation strategies respondents use with the three principal types of expenditure questions in the CE interview.

- Research with a small, heterogeneous, sample of respondents brought to a central location or interviewed at home would be appropriate. This research would obtain descriptions of the strategies respondents use to answer each of the types of questions for a variety of articles using concurrent think-aloud interviewing. The results would provide an indication of the variation among respondents in these strategies. The accuracy of answers obtained in different ways could be assessed using an appropriate criterion. Procedures would be developed and questions would be rewritten to encourage using effective strategies (such as remembering from recent to distant events or vice versa) (Loftus, 1984).

In addition, the way in which respondents interpret such questions as "What was the gross amount of ...'s last pay...?" should be investigated, both in focused interviews

and in the final instrument. Frequently such qualifiers are ignored: in making judgments about vignettes, for example, respondents have been insensitive to whether before-tax or after-tax income was specified (Schaeffer, Garfinkel, and Corbett 1986).

Aided recall

In their current form, the lists of items handed to respondents to aid recall are overcrowded, difficult to skim effectively because they lack visual focus, and visually fatiguing for both these reasons. The lists do not invite close examination: if the rapport of the interview suggests that speed is desirable, the temptation to give the lists an unfocused glance must be quite strong.⁵ These lists should be more attractive, provide a clearer focus, and be easier to read than they are now. Even lists that are more appealing will not be read or even skimmed unless interviewers encourage respondents to read each list at their own pace.

Conditioning Effects

Conditioning effects, as indicated by mean expenditures, occur somewhat erratically: speculations about the sources of these conditioning effects are necessarily somewhat ad hoc. Explanations include, for example, "reimbursements" for insurance, "shifting forward" for transportation, and "last effort" for gasoline (see *Conditioning and Recall Effects in the Interviewer Survey*).

- The criteria that represent interview conditions can be used to identify possible sources of conditioning effects: changes in the respondent within the household during the panel, changes in the time per item, changes in the use of filter items, and so on. Beginning after the initial interview, the list of items can be used to identify the type of memory errors involved and the type of items affected by each.

⁵I found it irresistible by the end of the interview.

CONSUMER EXPENDITURE DIARY SURVEY

The Diary Operational Test incorporates several promising instrument format changes: decreasing the size of the diary, decreasing the information requested for each item, and specifying the items on entry lines. Comments on most aspects of the diary format are best left until this experiment has been analyzed, and so my suggestions about the CE Diary are brief.

- Analysis of the diary data itself can be supplemented by in-depth focused interviews with a small heterogeneous sample of CE respondents after their participation in the survey about their problems in keeping the diaries. Such discussions should include both more and less cooperative respondents. The discussions should attempt to distinguish problems of motivation, problems of diary format and presentation, and problems due to using proxy respondents, because the solutions to these problems differ. In addition, observational studies of a small number of respondents completing the diary in different formats may suggest improvements.

Further analysis along the lines begun in *Reporting Quality in the Diary Survey* may indicate the need for new field procedures. It appears that the Diary survey obtains three kinds of data: data from records, data recorded by recall the same day, and data from longer recall periods. These three kinds of data differ in social and cognitive sources of inaccuracy. Inaccuracy in data obtained from records (such as receipts) will be due largely to transcription errors and incompleteness in the records; procedures to increase the accuracy of such reports might focus on identifying bias in the records (for example, are receipts missing for particular household members). Inaccuracy in data recorded at the time of purchase or at the end of the day should come from similar sources. Inaccuracy in data for longer recall periods, however, will be a result of more complex processes; procedures to increase the accuracy of such reports would focus on determining knowledge and aiding recall.

For example, if analysis indicates that reliance on recall increases during the two-week period, data collection in the second week should include a recall section. In addition, respondents who rely on recall for the first week can be identified and given special attention. Procedures are needed to help interviewers identify the reason the respondent relied on recall (for example, difficulty in writing) and find a solution (for example, calling during the week and recording the information over the telephone). Such intensive procedures to improve reporting may be feasible only if costs are decreased elsewhere. The advantages and disadvantages of adopting a one-week diary period should be studied.

Some suggestions made earlier about the CE Interview, are applicable here as well: motivating respondents may present particular problems with the diary survey and the use of financial incentives bears reexamination; developing respondent rules deserves close study here as well, given the large demands of the diary.

The Household Characteristics Questionnaire Supplement provides information valuable in analyzing performance on the diary. Some small changes may improve it as a tool for this kind of analysis. For example, recording the number of days on which recall was used and whether recall was used for all household members or only some would make possible analyses that would be useful in designing the field procedures suggested above.

CURRENT POPULATION SURVEY

The Report of the BLS-Census Bureau Questionnaire Design Task Force clearly outlines important conceptual issues and questions of understanding and cognitive processes that affect the Current Population Survey (CPS). The conceptual issues, however, are of primary importance. The information desired must be clearly defined before one can evaluate whether or not a given set of questions succeeds in obtaining it. If respondents are not able to provide the information needed to implement the definition required for analysis, it may be necessary to revise concepts to reflect the way in which categories are socially and cognitively organized. In the meantime, however, if the analytic definition of a "job," for example, remains ambiguous, research on the way in which respondents provide answers and interviewers code them will lack clear criteria for evaluation.

Many of the suggestions made in the discussion of the Consumer Expenditure Interview apply here as well, although with minor modifications. I will not repeat those discussions. The paragraphs below explain how the earlier suggestions can be extended for evaluating the CPS. (Words and phrases in boldface refer to sections or subsections in the discussion of the CE Interview.)

- **Motivating and training** respondents and developing a **rapport** that contributes to accurate reporting are as desirable for CPS as for CE respondents, and similar techniques are applicable. Increasing response accuracy is likely to require establishing a rapport in which the respondent feels comfortable taking the time needed to answer; this is particularly true for reports of actual hours worked.
- **Record use** is less applicable for most CPS questions about hours and income than for CE questions about purchases. However, increased accuracy and completeness of reports might be obtained if, for interviews after the initial interview, respondents were mailed a **self-administered worksheet** in advance of the interview, whether the interview was to be conducted in person or by telephone. The worksheet would

enable respondents who could use the form to obtain from all household members reports of (1) hours worked for all jobs for the reference week on a day-by-day basis, (2) earnings for the week from all jobs, and (3) job-seeking activities, before the interview took place.

- **Recall aids** such as calendars may improve the accuracy of reports about the hours worked in the previous week. The recall strategies respondents currently use and the accuracy of these strategies can be evaluated in comparison to the results the results obtained when the respondent fills in a calendar with number of hours worked each day (or starting and stopping time each day).

Criteria for Evaluating Changes

The criteria available for evaluating the effects of changes in the CPS interview are more limited than those available for the CE interview. This is true because presumptions about whether most errors result from overreporting or underreporting are even less clear for the CPS than for the CE and because the CPS requests less recall information than the CE interview. Still, a few criteria are available to use in assessing response accuracy:

- consulting others.
- response time or time to complete sections of the questionnaire.
- respondent is householder or spouse.

The comments made about these criteria in the discussion of the CE interview apply here also, and the usefulness of the criteria for the CPS would need to be evaluated.

- The possibilities for validating reports of hours and earnings are limited, and the few records respondents typically acquire (e.g., pay check stub) would not be available at the time of interview. Nevertheless, the feasibility of conducting validity studies should be investigated. For example, experimental treatments in which each worker

in a household kept a diary of hours and earnings for one week could be compared with the interview method. Alternatively, the interviewer could ask the respondent to save pay records when they become available. The interviewer would return to collect validating information about earnings and hours worked. It might be necessary to use self-administered forms to collect such information. Also, respondents in some occupations would not have records for all earnings (for example, waiters) or for all hours (for example, most salaried workers). Despite its limitations, such validations could suggest the limitations of accuracy in other reports.

Respondent Rules

Accurate information about employment, hours worked, earnings, and job search activities are probably not uniformly distributed within households. The objective of this research would be to develop respondent rules to identify the “best” respondent.

- Interviews with a relatively small heterogeneous sample can investigate both variation across households in the distribution of this information and whether or not household members agree about who is the most knowledgeable household member. Ideally, one would want to compare the proxy reports about each household member provided by (1) the most convenient household member to interview and (2) the household member judged “most knowledgeable,” with (3) members’ reports about themselves. Analysis should carefully identify information that appears unlikely to be known by any proxy, such as some kinds of job-search activities.

Job-Search Activities

Among other recommendations, the task force report suggests research about phrases to use in referring to job-search activities and the kinds of activities included in each phrase (p. 30). In addition to these investigations, the use of checklists or closed questions should be investigated. The principal reasons to avoid closed questions for a topic like job-search

activities are that social desirability pressures may increase mentions, particularly of relatively ambiguous categories, and the time involved in administering the question. Closed questions, on the other hand, leave less room for interpretation by the interviewer (which is an important source of potential bias in coding job-search activities) and present categories that serve as memory aids to the respondent. Furthermore, the categories complete the definition of the question by denotation and clarify for the respondent which activities are meant to be included in answers to the question. When categories are presented in checklist form (in which the respondent must reply "yes" or "no" to each item in the checklist) higher reports of most checklist items can be expected as compared to reports obtained when the interviewer shows a card (see, for example, Jordan, Marcus, and Reeder 1980). The factors that increase reporting would affect checklists presented in both personal and telephone interviews.

The checklist form of a closed question, however, has two additional advantages over other forms of the same question: the question can be almost identical for personal and telephone interviews and the checklist form may counteract the fast pace of telephone interviews and the tendency to give fewer mentions in response to open-ended questions in telephone as compared to personal interviews. This tendency, reported in Groves (1978), may help account for the decrease in the number of unemployed in the telephone interviews as compared to personal interviews.

- Whether or not these advantages are outweighed by the possibility of increased reporting of activities due to social desirability pressures or to the increased cost of administering checklists can be assessed by comparing results obtained with different question forms in both interviewing modes.

Discouraged Workers

The accuracy of the label "discouraged worker" depends in part on the reliability of reports of past job-search activities. The recall of past reports of job-search activities can

be evaluated to some extent using information gathered in previous interviews. In most cases the proposed question (Q23) about whether or not the person has looked for work in the past twelve months could not be validated directly in the current CPS design. An example of an exception is a person who says “no” in interview 5 but who reported searches in any of interviews 2 through 4. Special studies, however, can evaluate the accuracy of recall for shorter time periods.

- The recall of past job-search activities can be examined for up to a 4 month recall under the current CPS design. Longer recall periods can be used in interviews 5 through 8 if reports are restricted to the period covered by interviews 1 through 4. Whether or not the previous respondent or another was selected would be controlled.

Variation in Work Patterns and Earnings

Intensive interviewing can be used to determine how respondents decide whether or not they have regular hours (that is, how much variation can a work week have and still be considered “regular”) and how respondents determine typical or usual work hours (and earnings) and extra or overtime hours (and earnings).

- Concurrent think-aloud interviews can be used to identify the strategies that respondents use to make decisions about whether or not their work schedules are regular and about what their typical schedules are. The results of these interviews can be used to develop more structured interviews with a small heterogeneous sample of respondents to determine how much variation there is in these strategies. Validation studies using diaries would be an important supplement to these studies of reporting strategies.

Miscellaneous Comments on Proposed Test Questionnaire

It is not clear how “finished” this test questionnaire is or what its future will be. The questionnaire, however, is useful as a framework for commenting on some issues in the CPS, and I have tried to keep the following comments fairly general.

Q19 The introduction to this section should establish a frame of reference in such a way that respondents can use it. The effectiveness of the method used here and the use of aids such as calendars can be explored in interviews with a small number of respondents. In addition to establishing a frame of reference, the introduction should discuss the importance of accurate responses.

Q20 I do not understand why this question (or the one that it replaces) does not specify "for pay," which would clarify the intent of the question. Followup questions for those who answer "no" would then identify those who were pursuing business during the week but were not paid for their work. In the current version, "...any work *at all*" sounds challenging and is ambiguous.

Are all "no" answers probed to find out if the person is disabled? (An explicit followup question would be preferable.)

Q20B The report discusses the ambiguity of the phrase "on layoff" and it remains ambiguous here. If laboratory work does not produce a better phrase, all "yes" answers to a question like Q20B should be probed (preferably in a followup question) to make sure that the person expects to return to work.

Q20C If a person answers "no" to Q20 and "yes" to Q20B, there may be no code for them in Q20C (because jobs from which the R is on layoff are excluded).

Q20E Unless laboratory work suggests an alternative, use a filter question to determine whether or not the person has regular work hours. (How much variation there is in "regular" hours can be studied using small samples.)

Q20E-1 What does "usually" mean to people? How do they compute or construct this average or typical value?

Q20F Does this question intend to obtain multiple reasons or a "main" reason?

- Q20G What is the comparison that R uses to determine overtime or "extra" hours? Does this require computing an average for comparison or are they hours for which overtime pay is received?
- Q20K "Civic or military" which appears in Q21 might be useful here also. Why is "own illness" treated differently in Q20K and Q21?
- Q22A The report suggests that interviewers may vary in the way they interpret and code job-search activities (pp. 13-15). Problems of censoring by interviewers may be reduced by providing a list of activities that is comprehensive and specifying each activity as unambiguously as possible. Skip patterns and analysis programs can distinguish between activities which are considered valid job-searches and those that are not.
- Q22C The report notes on p. 23 that respondents who were previously employed may reasonably think that the question refers to the length of time since the previous job ended. For such respondents, the fact that a distinction is required may be communicated by first asking when the previous job ended and then asking during how many weeks since then the person has searched for work.

REFERENCES

- DeMaio, Theresa. 1980. "Refusals: Who, Where, and Why?" *Public Opinion Quarterly* 44:223-233.
- Cannell, Charles F., Peter V. Miller, and Lois Oksenberg. 1981. "Research on Interviewing Technique." Pp. 389-437 in Samuel Leinhardt, ed., *Sociological Methodology 1981*. San Francisco: Jossey-Bass.
- Groves, Robert. 1978. "On the Mode of Administering a Questionnaire and Responses to Open-Ended Items." *Social Science Research* 7:257-71.
- Lawrence A. Jordan, Alfred C. Marcus, and Leo G. Reeder. 1980. "Response Styles in Telephone and Household Interviewing: A Field Experiment." *Public Opinion Quarterly* 44:210-222.
- Loftus, Elizabeth. 1984. "Protocol Analysis of Response to Survey Recall Questions." Pp. 61-64 in Thomas B. Jabine, Miron L. Straf, Judith M. Tanur, and Roger Tourangeau, eds., *Cognitive Aspects of Survey Methodology*. Washington, D. C.: National Academy Press.
- Stinchcombe, Arthur, Calvin Jones, and Paul Sheatsley. 1981. "Nonresponse Bias for Attitude Questions." *Public Opinion Quarterly* 45:359-375.
- Schaeffer, Nora Cate, Irwin Garfinkel, and Thomas Corbett. 1986. "Respondent's Judgments about Components of the Child Support Reform." Unpublished manuscript. University of Wisconsin, Madison.