DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

National Center for Health Statistics
3700 East-West Highway
Hyattsville MD 20782

February 11, 1987

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

Dear Cathy:

Please find enclosed my report on possible research topics for CES and CPS. Although I am very busy on NCHS activities, I will be happy to find a little time to do the following, if this would be of use to BLS staff:

-- provide more detail about any of my proposals that you are interested in, and comment on the cognitive issues in any other proposals that are made.

-- suggest names of psychologists who might conduct short projects involving cognitive research.

-- offer practical advice and suggestions about the conducting of interviews using think-aloud techniques, probes etc., to test questions in a laboratory setting.

Thank you for inviting me to the Questionnaire Design Advisory Conference. I enjoyed the conference, and learned much from it. I was particularly pleased to find such interest in cognitive research amongst BLS staff and the advisors at the conference.

Sincerely yours,

David Mingay

David Mingay, Ph.D.
Service Fellow, Office
of Research and Methodology

Report on the Consumer Expenditure Survey

by David J. Mingay, Ph.D., Cognitive Psychologist, Office of Research and Methodology, National Center for Health Statistics.

Laboratory Research

Think aloud interviews and the cognitive labortory techniques mentioned by the BLS-Census Bureau Task Force on the Consumer Expenditure Survey are a powerful and fast way of learning of some of the difficulties respondents have with questions. In the absence of a special laboratory, an office can be used. A tape recorder to record interviews is all the equipment necessary, although the video taping of a few interviews would be useful. Some subjects might be recruited using standard CES procedures, and interviewed in their homes to avoid the problem of volunteers for lab studies being more highly motivated. Each respondent should be administered only part of the interview, as the techniques add considerably to the time involved. These interviews are likely to suggest hypotheses to be formally tested, and indicate possible improvements in the questions and procedures. If particular types of respondents are thought to have special difficulties with the survey, they could be efficiently recruited by means of flyers or advertisements. And as well as conducting interviews that exclusively involve laboratory techniques, some interviews might use the techniques to complement the more traditional experiments or in-house interview. For example, retrospective protocols administered after the questionnaire has been answered can indicate problems respondents have with questions.

Many of the studies I suggest below do not need large numbers of respondents. Perhaps 50 to 60 subjects in each group would be adequate, with relatively little attention needing to be paid to the demographics of respondents. Those findings that prove particularly interesting might be replicated, still on a fairly small sample, but one that is representative of the population.

Diary

The new diaries seem very well laid out. The large number of examples should increase reporting. It might be worth checking in the laboratory whether "etc." is understood by respondents (for example,131), by using probes, ie., that they are meant to write in items that are in the same categories. The extent to which respondents do this might be investigated, and a greater stress placed on the need to do so if such reporting is inadequate. More generally, the research on the Census form being conducted by Betsy Martin could very nicely be applied to the diary, to get a feeling for the ways in which respondents are filling it in, and the problems they are having with the form.

Recall Procedures

Research might explore the most appropriate ways of eliciting recall from those respondents who have not filled in the diary. Two possible approaches are:

1. Going forward or backward in time, i.e., from day 1 to day 7, or vice versa. One of these approaches may prove the most effective, and research could look at this.

2. Within each day, preceding the checklists with some questions designed to remind respondents of those activities he or she may have done during the day that are likely to be associated with purchases. Examples include going shopping, eating out, going on trips with children, and paying bills.

Cueing Respondents

Even if respondents conscientiously complete the diary at the end of each day, he/she may forget certain activities associated with purchases, for example a lunch time snack in the work canteen, or a quick visit to a store after work. Some fairly general instructions to respondents to recall what they did during the day, and the suggestion that they ask other members of their household too, might reduce underreporting. The instructions might range from the general instruction to recall the major activities during the day, from waking up to the present time, to a list of activities that are often associated with purchases, with the suggestion that these be kept in mind when completing the form. An alternative approach might be to have a similar set of instructions at the end of each day's section, with space for respondents to write in purchases previously forgotten about, (incidentally, a space for this purpose might be useful anyway, as respondents might recall products after passing the appropriate category in the diary, perhaps because memory for an item is cued by the recall of an associated item, and be unwilling or unable to locate the correct category). They could be encouraged to write in any items they had previously forgotten. The general instruction that respondents should either correct a previous answer or write these newly remembered items at the end of the questionnaire might reduce underreporting.

Other Approaches to Improving Data Quality

Several techniques occur to me that might increase respondents' accuracy when completing the diary, or increase the proportion who complete it after each day, although the techniques have the disadvantage of increasing respondent burden.

1. A specially designed notepad and pen might be attached to respondents' wallets or purses, so that they can jot down the cost and nature of single purchases as they make them.

2. Short diaries, given to other members of household.

3. The interviewer might telephone the respondent in the evening on several occasions, to enquire whether he/she has completed the diary for that day.

4. Ask respondent to make a commitment to trying to give accurate responses when given the diary, similar to those used by the Institute for Survey Research, University of Michigan. Research there by Charlie Cannell has suggested that this improves data quality. It might also be worth considering asking respondents to sign a form confirming that they have tried to give the correct answers after having completed the diary for each day.

Questionnaires

A number of possible ways of improving response accuracy occur to me. Unfortunately, these procedures add somewhat to the time that the interview will take, so might be better done in the context of a redesigned, shorter survey. In addition, they are relatively ineffective for proxy responses, so procedures for getting more of the household memebers involved in reporting purchases would be a very useful complement to the appraoch.

1. Cueing Recall of Puchases.

As the major problem appears to be underreporting, certain cueing procedures might boost recall:
Cueing by Types of Purchase

Survey research and cognitive psychologists know that recognition is better than recall. Reporting of apparel might be improved if respondents are asked to look through their wardrobes and cupboards, while thinking whether any of the clothing was purchased within the reference period. Respondents might also be asked to glance into several rooms looking for recently bought items, perhaps prior to the interview.

Cueing by Thinking of Events Associated with Purchases

Mental cueing might be engaged in. Respondents could be asked to recall major shopping expeditions, to think about them carefully and to recall the purchases associated with these visits. Similarly, respondents might be cued by being reminded of events which often lead to numerous purchases, for example, a wedding, a holiday, or a son/daughter going to college for the first time. Using probes in our laboratory, I have found some evidence that a respondent may forget a whole class of events. Hence, somebody who forgets to recall a wedding when recalling purchases may forget the clothes purchased for it, the gifts, the plane journey, hotel bills, and other expenses associated with the wedding. A checklist of infrequent events that are associated with numerous purchases might be read to respondents at the beginning of the survey, and subsequently respondents might be occasionally reminded of the events they reported as they go through the survey (see also comments on using calendar).

3. Cueing by Listing Specific Purchases

Currently, if respondents report not having purchased a product type, for example, small household appliances, the checklist of small appliances is not read out. This is likely to lead to underreporting, both because respondents have to think of a large variety of products, and decide if they fall into the category, and because respondents realize that a no response will reduce the number of questions they are asked. Recognition is much better than recall, so if the checklist is read to them, underreporting should be reduced. This procedure adds considerably to the number of questions asked, but may be appropriate for those items which are known to be seriously underreported.

4. Cueing Using a Time Line

This technique can be used both to get purchases dated more accurately, and to cue respondents to keep in mind events associated with major and unusual purchasing, although I feel the former is a more important use.

Respondents could be shown a calendar which extends a little way beyond the longest reference period on the questionnaire. Respondents would be instructed write in events which are (a) of some importance and (b) likely to have been associated with a number of purchase. Interviewers could indicate to respondents the type of events that are appropriate, for example, weddings, children starting college, major illnesses in the family, holidays. (Research might look at which types of events prove most effective as benchmarks.) Respondents could write in these events at the appropriate date on the time line (see example over page), and keep it in front of them while answering the subsequent questions. An additional use of the time line would be for the interviewers to indicate the appropriate reference period, especially when a new set of questions refer to a different reference period. For example, the interviewer might ask "In the last 3 months, that is, since October 20, 1986 ....", pointing to that date on the calendar, and marking with a red tape the 3-month period. This procedure may be particularly useful when the reference period excludes the current month, although it seems to me to be desirable that as much as possible, the reference period should run to the day prior to the day of the interview.

5. Cueing by Person

When a single question asks respondents to think of purchases by all members of the family, it is easy for them to forget that a particular family member made a relevant purchase, even though they knew of the purchase. Such information is more likely to be volunteered if the respondent is asked about each family member in turn. Of course, person by person questioning would add dramatically to the time of the interview. Despite this, selective use of the approach for some seriously underreported items could be useful. A research project might investigate which items this technique might be relevant for, by following the traditional family-based questions with questions asking other household members to report their purchases, and comparing the effectiveness of the two approaches in picking up these items.

6. Confidence Ratings

Many respondents can report how confident they feel about the accuracy of some of their answers. In laboratory research using simple memory materials, confidence is usually associated with accuracy, such that the more confident the subject, the greater his/her likelihood of being correct, although the association is by no means perfect. If a means of validating a number of respondents' answers was devised, I think that it would be of considerable interest to look at whether confidence was associated with accuracy on these questions. Respondents might, for example, be asked whether they are certain, fairly sure, not very sure, or guessing that their utility bill was $80-$85, or that two radios were purchased. Better estimates of purchases may be obtained if only those respondents claiming certainty are used in some of the analyses. A further advantage of the technique is that respondents may be pleased that their frequent feelings of uncertainty in an answer are noted by the interviewer, and therfore be more motivated to give considered responses. The converse could be true, too. Being able to report their answers as being guesses, respondents may feel that they are permitted to quickly guess, rather than work hard at giving a goodanswer. Much would depend on how the task was presented. As confidence ratings add to the time that questions take, only a subset of the questions could probably be treated in this way.

Another possible indicator of response accuracy is the information as to whether the data for an item has been taken from records. This question might be particularly useful when there is little risk that other relevant receipts or bills have been missed, for example, utility bills, or monthly payments on cars. I think that this is important information to elicit, and also tends to stress to the interviewer and respondent the usefulness of records. The information could be recorded by the interviewer, so this would add little to the complexity of the task asked of respondents.

Order of Questions

At the meeting, Nancy Mathiowetz pointed out that respondents are asked why they purchased an item before being asked what it cost. The way in which this type of information is organized in memory influences the ease with which different types of information can be retrieved, and which item of information better cues the other. The issue of which question order, if any, leads to better accuracy, would be well worth investigating, ideally in a record checking study. It may be that some respondents better retrieve price information following date information whilst others do the reverse, in which case a flexible interviewing procedure would be helpful. And perhaps certain types of products would be better reported when a particular question order is used. Alternatively, one order may be consistently superior, both in terms of time to retrieve the information and accuracy of the answers.

Similar issues arise on the questions on educational expense (Section 16), with eight different questions regarding each educational expense being asked in a particular order, Section 17, with cost by the period being followed by cost per month, and Sections 3D and 3E on mortgage and ownership costs. The best order of questions when

information on two or more time periods is required could be investigated, too. For example, in Section 14, hospitalization and health insurance, respondents are instructed to first think about payment in the last 3 months and then about payments in the last month. And in Section 5, costs are asked about in a forward order. Psychological theories do not permit a strong prediction as to what is the best order in which to ask these questions, but do provide reasons to believe that some question orders will be better then others.

8. Pace

Long questionnaires tend to promote a very fast pace of interview, as both the interviewer and respondent is often keen that the interview should end as soon as possible. However, the experience of the Institute for Survey Research is that interviewers can be trained to go more slowly, and that this slows the speed with which respondents give an answer. When respondents think for longer before answering, they are likely to give better responses. I recommend that ISR's training procedures for slowing the pace of the interview are investigated. This is hard to investigate in a small scale study as interview conditions are atypical, but a split panel study could be used.

9. Payment

Payment of respondents should be considered. A research project on these lines needs to assess data quality as well as cooperation rates, unlike in a previous BLS study. The money involved should be more than an nominal amount, and should be given either before or after several interviews to encourage continued cooperation.

Proxies

BLS does not know who makes the best respondent for the questionnaires and diaries. Hence, the current procedure of letting any responsible household member answer the questions is reasonable. However, certain household members are likely to be better respondents than others, who the best person is depending on the nature of the household and the type of products being asked about. Research could investigate these differences by testing each adult member of a small number of households, perhaps assuming that higher reported purchases indicates more accurate recall. If consistent differences are found in the survey itself, particular household members could be sought out to answer all or part of the questionnaire. For example, husbands might be asked some questions on cars and household repairs, and the wives other questions. However, as knowledge, interests, and responsibilities differ across households, a better procedure might be to first ask who would be the best person to answer different types of questions, or who makes the purchases, and then make an effort to interview that person on these items.

Some items may be badly reported by any proxies, in which case small additional questionnaires or diaries administered to each family member might be considered. Research might also investigate whether the same or different respondent should ideally be administered the subsequent questionnaires. If particular household members were found to be

especially good respondents, interviewers could be asked to attempt to interview these people, but be permitted to interview others if they should be unwilling or unavailable.

11. Redesign of Questionnaire

As was stressed at the meeting, the questions should map onto the way respondents think about the topics. For example, Section 20 might be made more flexible. Currently it demands that respondents work out a usual monthly amount. This may be an easy concept for some respondents, for some purchases at least, but not for other respondents. This section might be designed so that respondents can answer using the reference period they like (for example, weekly, monthly, or 3-monthly amounts, with the figures being adjusted at the data analysis stage). In addition, as is clear on the CPS, the concept of "usual" is problematic. Some respondents would prefer to give a total figure. For example respondents who only buy alcoholic beverages at Christmas may be resistant to reporting a usual monthly amount, even if they could do the necessary arithmetic.

12. Research on Time Period

A rather radical redesign of the CES, and one which I have not thought through in any detail, would involve moving to an event-determined reference period. A major reason for the decline in the recall of events over time is that events that happened recently interfere with the recall of similar earlier items, a phenomenon psychologists call retroactive interference (RI). The most accurate data should therefore come from the recall of a single, most recent event. Respondents might be asked to provide just this information, however short or long the reference period.

The different reference periods across respondents might be an analyst's nightmare, so I offer this suggestion with trepidation. In addition, this questioning procedure would lose a lot of data on frequently purchased items. It might be considered as a supplement to a shorter, perhaps one month reference period, if that change in reference period is to be investigated. In order to get sufficient data on infrequently bought items, such as cars, respondents might be asked to report the last purchase they made, however long has elapsed. Much of this data is likely to be quite accurate.

Conclusions

I would be happy to discuss any of these suggested projects in more detail. Most could be examined in a small scale study. I feel that the highest priority projects are the time line, pace, proxies, and perhaps certain question ordering and cueing procedures. Changing time line and question ordering procedures have the important features that they should not add to the time the interview takes, and by being somewhat easier for the respondent may even reduce total time.

Current Population Survey

David J. Mingay, Cognitive Psychologist, Office of Research and Methodology, National Center for Health Statistics

Laboratory Research

Many of the issues surrounding this survey concern the way in which respondents comprehend questions, retrieve relevant information, and select a response. The laboratory techniques which the BLS-Census Bureau Questionnaire Design Task Force list in their report are very useful procedures to investigate these issues, providing much detail about problems with the questionnaire. The Task Force appears to suggest that the laboratory approach should follow other activities, perhaps because of funding and other practical considerations. If circumstances permit, laboratory research should take place before the research. Useful research using concurrent and retrospective think-alouds, probes and other techniques does not require many resources. Interviews could be conducted in office or home, by BLS staff, using just a tape recorder. Some interviews should be conducted over the telephone in case respondents use different strategies under these circumstances. The lab approach might be particularly informative if the type of respondents who have difficulty with some of the questions are interviewed, for example, those who work greatly varying hours each week. Such people might be efficiently identified through companies, employment agencies, or advertisements in trade journals. The issue of the effectiveness of proxies might be explored too, if respondents' partners agree to be interviewed. Some lab techniques might nicely complement more traditional experiments. For example, retrospective protocols administered after the questionnaire has been answered can indicate problems that respondents had with questions, and can be used following a traditional interview.

The lab should provide useful information on how respondents are interpreting ambiguous questions, for example, whether hours worked is interpreted as including a lunch break, and whether alternative questions appear to work better. Lab work can also indicate how respondents who work irregular hours define the term "usual", and how those not directly employed by a company define the term "business". Some ambiguous terms might be interpreted in sufficiently similar a way, and in a way close enough to the survey designer's intended meaning that new wording is unnecessary. Unfortunately, this is likely to be the exception rather than the rule.

Suggested Research Topics

There is likely to be a conditioning effect for the questions on the number of hours worked last week and usual hours worked. Lab techniques could look at the change in strategies likely when answering whichever of these questions comes second, as a result of answering the first one. A study in which respondents answers are validated using company records might be conducted. Three conditions should be run: the two orders and a separation of the two questions by several other questions. Self and proxies may behave differently, so both types of respondents should be interviewed.

The research proposed below could be conducted using small samples, perhaps 50 to 60 respondents in each condition.

Dating of the Term "Last Week"

As the BLS-Census Bureau Task Force notes, the term "last week" is ambiguous. The National Health Interview Survey uses a calendar to anchor last week. This should increase the likelihood that respondents recall the correct period, indicates when the week finishes (some respondents might include Sunday and/or Saturday, and others might not), and emphasizes that last week is really meant, and not a typical week. A small study could look at the effect of using a calendar, using record checks. This could usefully be supplemented by laboratory interviews to see how people are currently defining the term "last week". Only the face-to-face interviews could be consistently conducted in this way, although significant numbers of respondents who are sent a calendar with instructions to leave it by the telephone may do so.

If the planned State supplement is to be introduced and the 2-week reference period used, the introduction of such calendars should be investigated in a similar study. It is possible that an event that happened on the first day of the reference period might be used to bound the event, but I suspect that the reference period is too short for this to be useful.

The incoming people who are asked about the 1-month reference period might also benefit from a calendar, as might respondents to question 22A on the current questionnaire regarding the last 4 weeks.

Changing the Response Period

If a change of reference period to two consecutive 1-week periods is made, there is likely to be interference between the two weeks. The most recent week is likely to be recalled with greatest accuracy, this accuracy being particularly high if the question is not preceded by the question concerning the other week. The calendar mentioned above may reduce error, but in general the week recalled second is likely to be recalled as being more similar to the first week than in fact it is.

Making question wording more in line with Respondents' Knowledge

Questions 21F and 22C of the proposed test questionnnaire are good examples of how respondents need to have knowledge structured in a particular form in order to answer questions. Some respondents may readily know, for example, what date they were on layoff, but they may not be sure how many weeks have elapsed since. If such difficulties are revealed in laboratory research, questions should be devised to give respondents an alternative way of answering. This is particularly easy in the CATI system, but could be used in face-to-face interviews as well.

Conclusions

I would be happy to discuss any of these suggested projects in more detail. I would also be happy to talk about how we have used the laboratory in looking at respondents' interpretations of questions and the strategies they use in answering them. Similar techniques would prove very informative in understanding problems that respondents are having with questions on the CPS. The use of a calendar to aid dating of events seems to me to be a particularly promising research topic.