

March 12, 1987

To: Cathy Dippe  
From: Judy Lessler *Judy*  
Subject: BLS Conference Report

At long last my report. I enjoyed the conference and thinking about the two surveys. Thanks very much for inviting me.

I hope that BLS will develop some activities that could be conducted under contract since I am very interested in working in this area.

/

COMMENTS AND IDEAS FOR THE BLS RESEARCH AGENDA FOR THE  
CONSUMER EXPENDITURE SURVEY AND THE CURRENT POPULATION SURVEY

This report comments upon the current BLS research agenda for the Consumer Expenditure Survey (CES) and the Current Population Survey (CPS) and presents some ideas in response to the two questions raised by Cathy Dippo at the Questionnaire Design Advisory Conference held on January 15 and 16, 1987. Specifically, it considers:

1. What types of research should BLS be conducting; and
2. What are the priorities among the various research topics.

The first section of the report focuses on the CES.

#### 1.0 Consumer Expenditure Survey

##### 1.1 The Experience of Being Interviewed

Some of the ideas that are included in this report arise out of the experience of being interviewed. The interviewer came to my office early one morning. Except for a few telephone calls, there were few distractions during the interview so I was able to concentrate on the interview. The interviewer decided to conduct the interview that is received by a replacement household since she said that it was the most complicated and would give me a good idea of the most

difficult response task faced by a respondent. The interviewer was helpful, friendly, and professional. She had obviously thought about the problems of that respondents have and how good data could be collected.

In spite of the interviewer's skill, I found the response task to be enormous and had some difficulty determining what I was to report. Part of the large respondent burden was a function of the time period the interview was conducted. The period covered was from September through December 15, 1986. During this period our consumer unit (CU) had had school expenses for 5 children, major expenditures for clothing (which in our family are largely confined to 2 major shopping trips per year), the majority of the expenses for Christmas gifts, expenses for the Lessler Family Birthday Season, and expenses for a trip and fete in celebration of my parent's 50th wedding anniversary. In addition, we had expenses related to the destruction of our water pump, washing machine, and tractor barn during a large thunder storm.

In addition, to the recall problems inherent in having a large number of expenditures to report, it turns out that I was attempting to report living expenses incurred by the three children that were enrolled in college, and did not need to report these by category of expense. I was, and still am, unclear as to whether certain expenses are to be reported when they were incurred or when the bills are paid. I had difficulty determining in which month the expenses occurred and made heavy use of the 50th fete (which occurred on October

10) and Thanksgiving as landmark events.

I also did not know the cost associated with a large number of the expenditures since my husband pays a good number of the bills and I am unaware of his out-of-pocket expenses for food, telephone calls, haircuts, and so on. The interviewer indicated that if our CU was truly a member of the sample that she would call back to get the value of the expenses that I did not know.

I concluded that I was not a good respondent. I had difficulty recalling expenditures and their associated costs, difficulty deciding what to report, problems placing things in time (although I am pretty sure that I had them correctly placed in relation to the two landmark events), and did not have knowledge of some events. I also got tired and ended the interview early--some 3.5 hours after beginning.

Not every respondent will experience this difficulty. If I had been interviewed on August 15, I, myself, would have had a much easier response task as there would have been no school, clothing, party, or Christmas expenses to report. However, my difficulty has important implications to the survey. When there is a great deal to report there are greater difficulties and greater error in the reports. The quality of data for specific times of the year may vary.

One interesting thing has occurred since I participated in the survey which might have some implication to the decline in reports seen over the reporting period. Our CU changed its behavior. We reasoned that we did not have a

good idea of how or a plan for the way we were spending our money. We have started monitoring our expenditures and are spending less--particularly for small miscellaneous items. Do others do this? Could this contribute to the decline in expenditures over the reporting periods? This could be investigated by having the interviewers ask the respondents whether they had changed their behavior since the previous interview.

*Change  
behavior*

## 1.2 Current BLS/CES Research and Future Plans

This section presents a few thoughts in reaction to the research that has been conducted and plans for future research. It is apparent that BLS is concerned with the quality of the data and committed to improvements. The experiments conducted by C. Tucker in connection with the diary component are particularly innovative. The idea of having measures of response quality built into the survey and using them to make adjustments in the survey estimates is being investigated in a variety of settings and is gaining acceptance. The scales that are created from the diary have the potential to form the basis for such an adjustment procedure. Some of the experiments could be more valuable if there was a means for assessing the true number of and values of expenditures, although, this is of course difficult to do.

I have some reservations about the field experiment planned for the redesign. If I understand the experiment

correctly, it basically repeats the the experiment conducted by Neter and Waksberg (1964) with the addition of a randomization of the sections to measure fatigue effects. While the plans are sound, the I think that we might already know what the outcome is. Reduced respondent burden and reduced recall periods will seem to be related to better reporting. Given that these are likely results, what decisions will be made based upon these results? It might be better to go ahead and some make these decisions now and focus on the what changes need to be made in the survey to meet the goals of better data. Presumably this would consist of identifying the best data collection protocol for each type of item considering the accuracy that is needed to met the analytic objectives of the survey. The next section discusses an approach to the redesign.

### 1.3 Assessment of the Overall Survey Design

Much of this discussion is based on the following assumptions and beliefs:

1. The response task inherent to the CES is enormous and a large number of respondents have difficulty with the task.
2. That many reserachers who use survey data have never faced up to the difficulty of obtaining survey measurements by asking questions. This is not true of just BLS but of other (all?) surveys. For example, one researcher wanted to find about the psycho-social functioning of certain family members and asked the family informant "How would you rate the psycho-social functioning of \_\_\_\_\_?"
3. That survey design (and survey modifications)

6

should always be done from a total survey design approach which means keeping in mind the effect on total error of decisions that are made about individual components.

Based upon this orientation, my ideas as to the research agenda for the CES are as follows:

REEVALUATE THE SCOPE OF THE SURVEY. The entire set of characteristics needs to be examined to determine if it can be reduced in some way. Some questions need to be answered, such as:

Who are the end users of the data? Do they need the entire complex of information that is currently contained in the interview?

Are there better or alternate ways of meeting the analysis goals of some of the users other than collecting interview data?

What are the required levels of accuracy for each of the information items?

Are these levels currently being met?

The answers to these questions may provide some insights into ways in which the scope of the interview could be reduced.

FORMS APPRASIAL. Conduct a "forms appraisal" to determine where to direct the research efforts. Consider what types of errors can occur and the seriousness of each in terms of meeting the analytic objectives of the survey. For example, is failure to recall an item entirely more serious than failure to correctly place it in time or to recall the cost of the item. Alternately reporting items that should not be reported may be serious since it can produce overestimates.

reports  
y  
reports

+

*(Consumers)*  
*speculate*

An example is when the respondent is confused as to whether certain costs are associated with his or her consumer unit. During the forms appraisal one should consider variation by type of item, type of respondent, difficulty of recall, self versus proxy response, and the balance between the recall survey and the diary survey. One should speculate on the effect of these differences on the results.

ADAPT MEASUREMENTS TO RESPONDENTS. Consider how the measurement process can be adapted to the respondent and the structure of the CU. The needs and competency of the respondents need to be balanced against the goals of the researchers. Other ways of getting the data should be considered. For example, could some items be gathered by merely noting the fact of purchase and others include the cost and time at which purchased. Some ideas on this are presented in the section 1.4. Currently the burden of making the measurements needed for the survey is placed upon the respondent. He or she has to recall the purchase, place it in time, allocated it to persons within the CU, and report the cost. This requires monitoring one's behavior in ways that are unusual.

Just consider what is necessary to provide good data for the telephone expenses question. At no cost to the government, one must retain each telephone bill for up to 3 months, must note the month in which the bill was received which may be different from the month issued, must, upon arrival of the interviewer, examine each bill to determine the



total basic service charges which may entail adding across several items, totals for taxes and long distance calls, calculate the amount that will be deducted as a business expense and calculate the percentage. This is a large departure from opening up the telephone bill, scanning it for errors, writing the check, tossing the bill and enclosed flyers into the trash, and telling the children or spouse that their long distance charges are out of line.

Books and entertainment expenses is another item which adequate response would require a considerable change in behavior. In our family, and in a lot of others I suspect, books, magazines, movies, and so on, are "pocket-money" expenses--if you have some money in your pocket you buy a book, paper, or magazine. This continues, for the children, until the funds have all been spent. Whether or not one buys the items on display at the checkout counter is dependent upon the length of the line and what looks interesting. That is why they put them there. What family can exercise enough control so that each child will keep a 3 month tally of his or her expenses for such and report it to the family informant in time for the interview.

Many other items similarly require noting and monitoring things that one does not usually monitor. Thus, the respondent is required to make the measurements for the researcher performing tasks such as retaining records, making tallies by category, making allocations across time by category and (sometimes) by person, and calculating amounts and percentages. The interviewer asks about each item; the

respondent does everything else.

I think that the entire complex of activities that a CU would be required to complete in order to give accurate data should be sketched out. This, I believe, will reveal that, for many items, at best approximate answers or general impressions are being collected. And the degree to which data are approximate will vary greatly from CU to CU. Variation across CUs in the quality of the information may be a more serious problem than just some level of inaccurate reporting that tends to affect all CUs in a similar manner. This is because of the detrimental impact that such variation will have on relational analyses.

After considering this complex of activities that are required to be a "good" respondent, the researchers should consider modifying the measurement process. Maybe accurate data are not needed for each item. Modification of the measurement process would entail discovering what respondents can reasonably be expected to do, what they can (and will) do well, and then designing the measurement process to accommodate to the abilities of the respondent. This will require experimentation, and a series of laboratory studies would be ideal for investigating the capabilities of respondents.

*Consider  
 the  
 possibility  
 of  
 a  
 different  
 measurement  
 process  
 to  
 meet  
 the  
 needs  
 of  
 the  
 respondent*

Redesigning the measurement process may require a different measurement method for different types of CUs. For example, in terms of the telephone items, I would much rather save my bill and hand it over to the interviewer to tally--or be given a form that I could fill out at the time that I

opened the bill so I would not have to save the bill and could do the recording at my convenience--or have the telephone company send a copy of the bill to BLS.

*former  
customer  
or  
to  
BLS*

A different measurement process for different CUs could entail different ways to collect the same data, and it also could entail having different subsets of data for different respondents. This would require an estimation method that accounts for these differences.

EMBED QUALITY MEASUREMENT IN THE SURVEY. The survey evaluation and redesign should be an ongoing effort.

Iterative testing procedures and methods for continually monitoring the results could be employed. For example, some portion of the sample (or a supplementary sample) could be designated for intensive study each year. A section of the interview (or diary) could be studied each year so that over some period all components have been studied. For example, one year BLS might want to focus on the respondents' understanding of the concepts and their general behavior, such as their ideas of a CU, the time period for reporting, what they think they should report, their reactions to the interview, their changes in behavior, and so on. Another year might be devoted to studying the section on utilities, another the daily diary and so on.

*part of  
sample  
or  
supplementary  
sample  
studied  
each  
year  
(K-H-S)*

*continuous  
3 year cycle*

Each component studied might involve a 3 year cycle. One year for planning and laboratory studies, one year for field studies, and one for analysis and modifications in the instruments. Each year, however, some component would be in

the field testing/study stage.

In addition, I think that BLS should begin to investigate methods for building quality scales into the interview that could be used to adjust the estimates and on the measurement process.

#### 1.4 Ideas for alternate measurement methods and experiments

This section presents some ideas of experiments that BLS could conduct and alternate measurement methods that could be considered.

INITIAL OBSERVATIONAL STUDIES. There are a series of interesting observational studies that could be conducted. Some respondents could be asked to provide response protocols (think-aloud interviews) for sections of the interview survey. This could be people who have never participated in the survey. In addition, visiting some people who have been in the survey and interviewing them about how they answered the questions would be interesting.

Some observations of diary respondents would be interesting. The alternate smaller diaries that were handed out at the January meeting have made major changes in the response task. It is easier in that the respondent is not required to list items; however, it is much harder in that respondents are required to group items by category and then add up the cost of the items in that category. This would

*But how  
to report  
on this  
in survey  
House*

require that the respondent have a work sheet to add the costs of all fresh vegetables, different types of paper products, and so on. This is a considerable increase in difficulty over the other version of the diary. Also some of the categories are not clear--they have etc. as a qualifier. What goes into the "etc."?

Respondents are likely to differ in their interpretation of what is a member of a category. For example, if one buys a number of loaves of bread and puts them in the freezer, are these frozen bakery products or bread nonrefrigerated? What are the "related products" for "ice cream, frozen yogurt, and related products"? Where does one put hushpuppie mix--under "flour and prepared flour mixes" or under "spaghetti and other pasta, cornmeal"?

It would be interesting to observe some people filling out the form and see how they decide what to do. Do they truly sum prices for all items or do they just list the price for one of the items in the category? How good are they at doing the sums? How do they decide to categorize items? Getting into the home to observe the recording of the items could be difficult. Several intermediate experiments might consist of bringing some people in and giving them a pile of food and other goods to record and observe the task. Or one could ask a group of respondents who have grocery store receipts that list the item and other types of receipts to bring these in and observe them completing the diary. They should be asked to tell what they were thinking while they completed the diary, and this should be filmed and/or

*different  
different  
different  
different*

*set market  
market*

recorded.

It is possible that some respondents would agree to have observations conducted in the home. The observer could make an appointment to come by at the time the respondent was recording the day's expenses. Another way of getting response protocols in the home would be to provide the respondent with a recorder and ask them to provide the protocol to the recorder. This would have to be tested first to see if it would work.

A good way of involving the analysts in the design of the questionnaires would be to have them provide response protocols for both the diary and the interview. This would give them insight into the types of thinking that go on during the response process.

RECALL EXPERIMENTS. The diary sample would provide a good vehicle for some recall experiments. A subset of people who had completed diaries could be visited at later times--1 week, 1 month, and 3 months--and the same information collected by recall. To increase the value of this experiment, BLS might want to design an experimental diary that included more of the items on the interview. Respondents who had recorded information in the diary could be expected to have better recall than the general respondent because the act of recording the purchase would make it more memorable; however, the results would place limits on the recall abilities of respondents. In addition, one could have several members of the household participate in the experiment. One

*Very good  
in recall  
experiments*

leading  
cognitive  
influence

member could do all of the recording and each could be interviewed at a later date. In addition, one could have the respondents rate their ability to recall items and see if that is related to the the actual recall accuracy.

Another interesting recall experiment would be to try and get some idea of how the recall task is best approached. Is it better to recall all items purchased before attempting to place the purchases in time, allocate them to persons, and recall their costs. A group of persons for whom the purchases were known could be asked to perform the task in different ways and the number of events recalled under each scenario could be determined. I found the item lists very helpful for cueing recall. It might be advantageous to give the respondent a checklist so that he or she could read down it checking off items that had been purchased. The think-aloud interviews would give some ideas as to how the respondents approached the recall task and how questioning strategies could be developed to assist in the recall process.

Another interesting experiment would be to investigate the error in recall of purchase amounts. Do respondents consistently over or under estimate the amounts spent on various items? What are the absolute values of the recall errors? It may be that for some items, the recall errors are so great that it would be better to focus on the fact of purchase and to gather costs of the item from some other source. The items in the survey differ according to the degree to which the amount paid for the item can vary from CU

to CU. It is not unreasonable to expect the cost of a dress to vary by factors of up to 100--ten dollars to 1000 dollars. However, for some other items, the cost might not vary so much, e.g., thread, children's polyester and cotton underwear, and so on. It might be that the recall variability is greater than the existing price variability or having the respondent indicate the type of outlet at which the purchase was made and then using some other source for the average costs of items would be more accurate than depending on respondent recall. This would allow the respondent to focus on the number of items purchased thereby simplifying the recall task. Also, merely indicating whether the cost was within a certain range might meet the analytic needs of the survey.

*considerable  
of them*

*Cost in  
certain  
range*

THE INTERVIEW PROCESS. There are a number of factors that have been shown to effect the quality of interviews. Charles Cannell has conducted a large number of studies in this area, and I hope he is mentioning them in this report. However, I will mention a few of these here.

Several of the participants in the conference indicated that the pace of the interview was excessive and that the interviewer seemed to be impatient with the amount of time that the respondent was devoting to attempting to recall the information. If respondents feel a preceived pressure to be quick, this can lead to a number of assumptions on their part as to what they should report. This may lead them to conclude that approximate answers are all that is needed and that only important things should be reported each giving his



or her own interpretation of what is important. Longer interviews might result in higher nonresponse and more breakoffs within the interview. However, it might be worthwhile to explicitly examine the tradeoffs between a fast paced interview and a more leisurely paced interview. A number of the experiments that were conducted at NCHS in connection with its pilot project on the use of the cognitive laboratory for the design and testing of questionnaires seemed to indicate that pace had an important effect on the results.

*fast vs. slow  
Pace*

*respondents' need  
commitment*

Another fruitful area of study is "commitment to give good information" and reinforcing, within the interview, behaviors that are likely to be related to high quality reporting. Cannell and his colleagues have conducted a number of studies that show that direct attempts to set goals for good reporting and reinforcement in the interview have an impact on the number of events reported. These effects have not always been the same across subgroups of people; however, there is enough evidence of positive impacts that such issues deserve exploration.

The use of proxy respondents and minimally cooperative respondents is another area that could be studied. One of the field staff supervisors who was attending the meeting commented that in some cases the interviewer was forced to conduct the interview while standing on the door step. I question whether, given the reporting task inherent in this survey, such an interview is worth much. One could expect that an interview conducted under these conditions

should cover only some key items that would allow one to do a good job of adjusting for the remaining missing data.

The interviewer's task needs to be simplified. My interviewer, who seemed to be experienced and professional, seemed to spend a lot of time in bookkeeping operations--recording here and there, copying information from one place to another, and so on. In addition, when there was some question as to whether or not I was to report some particular types of expenses, she would ask me a series of "diagnostic" questions the answers to which caused her to decide to include or not include some item. For example, we spent some time trying to decide whether the basement in our house was full or partial, whether I was to report for rental properties that were investments but in which our children who are enrolled in college are also resident, and whether the way in which the college children's living expenses are paid for constituted an expenses for our CU. There is nothing in the interview that records the decisions that were made as a result of these questions. Several of these decisions were critical to the type of data that were collected.

CHP I  
Given the complexity of the interviewing task, I think that BLS should seriously consider moving to Computer Assisted Personal Interviewing (CAPI) in which the interviewer is provided with a computer which she takes into the home. This computer could be programmed to take care of the bookkeeping tasks--skip patterns, checks on previous rounds of reports, and so on. This would free the interviewer to focus on the respondent/interviewer interaction and to assist the

respondent in the response task. In addition, CAPI would make it possible to implement some of the innovations in the measurement process that have been suggested. It would be easy to randomize the order of the sections, to build in different types of questions that are directly related to the particular type of CU that is being interviewed, to standardize and record the answers to some of the questions that the interviewer asks in order to determine if something should be reported. Questions on whether or not the respondent bought an airplane could be automatically skipped in some households.

The idea of using CAPI in the interview brings me to my last point and that is changing the design of the survey and the estimation procedures so that the quality of each interview is assessed and that this information is used in the production of statistics from the survey.

EMBEDDING QUALITY ASSESSMENT INTO THE SURVEY PROCESS. A number of studies have shown that response quality varies with characteristics of the interview and the respondent. Tucker's studies on the diary are an example. LaVange and Folsom (1985) reported on the use of regression estimates of the effect of interview characteristics to adjust for nonsampling errors in the National Crime Survey. They investigated reporting associated with forward telescoping, time lag between interview month and month of victimization, conditioning effects, proxy interviews, telephone interviews and household interviews. Methods were proposed for producing

model-based predicted victimization rates that were adjusted for the effects of the survey measurement conditions.

These types of estimation techniques combined with quality assessment techniques built into the interview have potential for improving the results from the CES. CAPI interview methods could be used to permit random assignment of alternate versions of the questionnaire across households. In addition, items needed for quality scales such as those that were developed by Tucker for the diary component, could be imbedded into the interview. Such scales could also be used to guide the interview process. For example, a very simple scale that assessed the interview conditions could be used to direct the interviewer to only ask certain questions. Are you conducting the interview while standing on the porch in the rain? Is the previously interviewed respondent available? Did the CU save any receipts? Have there been any major changes in the CU in terms of persons living there or living conditions? Answers to simple questions like these could be used to adjust the interview so that the most important and most accurate data could be collected given the survey interview conditions. Alternately, they could be used in the estimation process to adjust for the survey conditions.

*Interview  
condition  
scales*

Of course this could not be implemented tomorrow--or even next week. Determining what to include in the scales, how to adjust the interviews, and so on would and should be an ongoing process that is imbedded into the entire survey program. Research into recall, compliance, interviewer behavior, reference periods, question wordings, questionnaire

structure, etc. would all need to be ongoing. However, it seems that the time to start working on some really innovative approaches is now.

## 2.0 Current Population Survey

The report by the task force was excellent and very complete. I have been able to think of few comments, however, here are the few:

The schedule seems protracted. I think it would be worthwhile to begin with some group sessions and response protocols immediately.

The research should focus on unusual situations. Persons with a single job probably have few problems with the questionnaire. Focus on people who are likely to shift in their classification.

The principle of making all questions explicit is EXCELLENT. However, the first question of the revised questionnaire violates this principle since it allows for a response not included in the question.

I looked at the revised questions and tried to imagine several people I know answering them. I think several situations would slip by with the current questionnaire. These include a woman who runs a catering business from her home, a woman who keeps the books for her family farm and also works at a parttime job, and a woman who works full time at RTI and also runs a small farming business. I have developed a

revised set of beginning questions that I think would find these situations. They are as follows:

Alternate wordings:

20a. Does anyone in your household operate a farm or their own business?

Yes...01 [ASK 20b]  
No....02 [GO TO 21]

20b. Last week, did you do any work at all for this farm or this business?

Yes...01 [GO TO 21]  
No....02 [ASK 20c]

20c. Do you usually work do work for this farm or business?

Yes...01  
No....02

21a. Last week did you have a (another) job?

Yes...01 [ASK 21b]  
No....02 [GO TO 23]

21b. Did you have more than one job last week. This could be a part-time job or a weekend job? Do not count jobs from which you are on layoff.

Yes...01 [ASK 21c]  
No....02 [GO TO 22]

21c. How many jobs did you have?

2 3 4+

22. LAST WEEK, did you do any work at all at your job (any of your jobs)?

Yes...01  
No....02

23. INTERVIEWER CHECK ITEM

USUALLY WORKS AT A FARM OR FAMILY BUSINESS...Yes....No  
WORKED LAST WEEK IN THIS FARM OR BUSINESS...Yes....No

HAD AT LEAST ONE JOB.....Yes....No  
HAD MORE THAN ONE JOB.....Yes....No

WORKED LAST WEEK AT A JOB.....Yes...No

VERIFY INFORMATION SAYING:

EXAMPLES:

I have recorded that you worked at a job last week and that you usually work at a farm or family business. Is this correct?

I have recorded that last week you worked at a job and at a family business. Is this correct?

I have recorded that you did not work last week but that you have a job. Is this correct?

I have recorded that you did not work last week but that you usually work in a farm or family business. Is this correct?

I have recorded that you did not work last week and that you do not have a job. Is this correct?

AT THIS POINT THERE WOULD BE 5 SETS OF QUESTIONS THAT PEOPLE WOULD BE ROUTED TO ACCORDING TO THE CLASSIFICATION ABOVE:

SET 1: NO JOB, FARM OR BUSINESS

SET 2: WORKED LAST WEEK--ONE JOB (OR ONE FARM OR BUSINESS)

SET 3: NO WORK LAST WEEK--ONE JOB (OR ONE FARM OR BUSINESS)

SET 4: WORKED LAST WEEK--MORE THAN ONE JOB (OR AT LEAST ONE JOB AND A FARM OR BUSINESS)

SET 5: NO WORK LAST--MORE THAN ONE JOB (OR AT LEAST ONE JOB AND A FARM OR BUSINESS)

It would be easy to implement this on the CATI system or on a CAPI system.

## REFERENCES

Neter, John and Joseph Waksberg (1964) "A Study of Response Errors in Expenditures Data from Household Interviews," JASA, Vol. 59, pp 18-55.

LaVange, Lisa M. and Ralph E. Folsom (1985) "Regression Estimates of Survey Operations Effects: Adjustments for Nonsampling Bias," Proceeding of the ASA 1985.