

Cognitive Design and Bayesian Modeling of a Census Survey of Income Recall

**Kent H. Marquis (US Census Bureau) and S. James Press (University of California, Riverside)
with the Assistance of Meredith Lee (US Census Bureau)**

This is a progress report on combining new thinking about Bayesian estimation and cognitive psychology to the problem of making estimates using data that may contain response errors. It is joint research inspired by Jim Press while he was at the Census Bureau as an NSF/ASA Fellow in 1997-98.

The basic problem is how to improve population estimates from surveys or censuses when the responses contain response errors and the distribution characteristics of the response errors, say the first two moments, are initially unknown. Jim's basic idea is to elicit auxiliary information from respondents about the quality of their response. And to use it in an empirical Bayes estimate of the population parameter which would be more accurate and have less variability than a traditional parameter estimate.

Jim had conducted two pretests on college campuses that showed promising results, except for the tendency of some students to omit answers or give unacceptable answers, perhaps due to a lack of understanding of the auxiliary information task. In his research proposal for his fellowship, Jim asked for collaboration with the Census Bureau's cognitive scientists on constructing an understandable task for respondents. This paper reports on that collaboration.

At the Census Bureau, we are interested in the general issue of how well respondents can judge the quality of their answers. If respondents can judge well, Jim's approach might be quite useful, and, if it didn't work out, there might be other ways of using accurate auxiliary information to improve estimates that were otherwise unadjusted for response errors.

The ability of respondents to know the quality of their answers is an instance of Metacognition, an emerging field that is beginning to attract both theoretical and applied attention within the general area of cognitive psychology. Our desire to bring this body of theory and research to bear on the problems of questionnaire measurement is what motivated the Census Bureau's support of the extension of Jim's research.

This paper will discuss the auxiliary information part of the project. We will cover the use of the information to improve estimation in a future paper. First, we will discuss metacognition. Then we will describe a series of three cognitive research studies conducted at the Census Bureau to learn how to formulate workable questions about metacognition. Third, we will describe a larger scale telephone survey that we conducted to test the revised questions and some preliminary results from that survey. On the basis of the available data, we will show that, while we have not solved all the measurement problems, the data appear to contain enough additional information to be useful in improving our estimates.

2. METACOGNITION

Metacognition refers to what we know about what we know (see, for example, Metcalf and Shimamura, 1994).

When we encounter a question, metacognitive theory says we have a feeling of knowing (FOK) about the answer. If we think we know the answer, then we will go ahead and work on the task, eventually arriving at an answer that we report.

Sometimes, however, we have a tip-of-the-tongue (TOT) experience, where we are sure that we know the answer but we just can't retrieve it at that moment. In this case, the metacognition is that we know that we know the answer but are just having retrieval problems.

In other contexts, according to the theory, we constantly make judgments about how well we have mastered the learning of some body of material (JOL). Based on those judgments, we decide how much attention and effort to give to learning more and we decide to what areas we want to devote more of our resources. We monitor our learning progress,

to judge when to stop. In our research, we address a similar concept, the metacognitive judgments of answer accuracy (JOA).

Current theoretical work addresses how we make judgments of knowing. Some proposed mechanisms postulate that we are capable of making very accurate metacognitive judgments. Other hypothesized processes suggest that our metacognitive judgments can be very biased and inaccurate (e.g., Metcalf, Schwartz and Joaquim, 1993; Koriat, 1994; Reder and Schunn, 1996).

Our general research issue about questionnaires, then, is what do we know about the quality of the answers we give. More specifically, when we construct an answer to a factual question by retrieving information from memory, can we accurately judge the goodness of those memories and hence accurately infer the quality of the answer we constructed?

The applied metacognition literature suggests that we have some information about what we know and how well we know it, but that such information is not completely accurate. One growing body of literature concerns eyewitness testimony (e.g., Sporer et al, 1995). A typical experimental arrangement is to show subjects a video clip of a crime being committed, ask them questions about what they witnessed, and to rate their confidence in their answers. Correlations between the correctness of the answers and the confidence ratings generally are above the chance level but far below perfect values, generally falling in the .60 to .80 range on, say, multiple-choice answering tasks. In general, subjects tend to show an overconfidence bias. But recalibrating the data for such biases does not necessarily increase the correlations.

In the cognitive laboratory studies and telephone survey, we asked questions (a) that cannot usually be answered by retrieving a single element from memory (where the answer must be constructed) and (b) that are difficult enough to result in metacognitive judgments of not knowing. For the telephone survey, we obtained external criterion or truth data to learn how accurate the metacognitive information is.

3. DEVELOPING QUESTIONS IN THE COGNITIVE LABORATORY

Our initial goal was to develop questioning procedures to elicit the standard answer and the range of plausible alternative values. For estimation purposes, we wanted to get quantitative, interval scale information useful in fitting a Bayesian prior distribution for each respondent. So we decided to ask about income. To cover a range of difficulty, we asked about two types of income for the most recent calendar year (1997) and the year before that. Then asked how much each of the two types of income changed over the past five years. The income types were wages and salaries on the one hand and interest and dividends on the other.

Jim brainstormed many different ways of obtaining the main and auxiliary information. We submitted the brainstorming results to cognitive experts at Census who screened out some of the more impractical or incomprehensible ideas. Then we tested the remaining approaches in three rounds of cognitive interview studies.

3.1 First Study - Our purposes, in the first laboratory study, were to:

- Test respondents' capacity to understand and answer the basic questions.
- Test their ability to comprehend and perform the range-definition task.
- Test different ways of asking the range questions.
- Test the order of asking the standard and range questions: (e.g., standard question before or after the range question).

3.1.1. Methods - We interviewed 10 respondents individually in our cognitive laboratory, by simulating a telephone interview. The respondents, as a group, were in the low and middle family income range, married and living with their spouse, and worked for wages or salaries within the past five years. We interviewed blacks and whites, males and females, younger and older persons. All respondents were paid for their participation.

We tried several kinds of questions and tried different wordings within the question types. Here are some example questions that we tried:

Standard: How much was your total household income from salaries or wages in 1997?

Range Example 1:

Please give me two numbers. One that you're just about sure is smaller than your total household income from salaries or wages in 1997, and one that you're just about sure is larger than your total household income from salaries or wages in 1997.

Try to make the two numbers as close together as possible while still being sure that one is below the true value and one is above.

Range Example 2: Give me a number so that you would be very surprised if you found out that your total household income from salaries or wages in 1997 was LESS than that number [Analyst would then assume a symmetric interval and impute the highest value].

Here is an example item that asks for the range information first:

Now I am going to ask you some questions about your total household income from interest and dividends during 1997.

What is the smallest interval you can give me so that you believe that the true amount of your total household income from interest and dividends during 1997 will be in the middle of that interval?

For all interview sessions our procedure was to start with an icebreaker question that showed our interest in the respondent's well-being. This question also set the stage for the income questions:

We're interested in how people are getting along financially these days. Would you say that you and your family are better off or worse off, financially, than you were a year ago?

After that, we counterbalanced the order of standard and range questions within interviews. Different interviews contained different sets of questions so that we could test as many as possible.

During each interview, we asked respondents to think aloud as they thought about the questions and answers. The cognitive interviewer used probe questions as necessary to understand the respondent's cognitive processes. The interviewer used general probes (e.g., Can you tell me more about that?) and metacognitive probes (e.g., How did you tell that your answer was as correct as possible?)

3.1.2 Results - Study One revealed a large number of problems due to the questioning procedures. For example:

1. The auxiliary questions were too long for comprehension over the telephone. Respondents often asked that the question be repeated. Several key terms were not always understood, these included fundamental terms such as interval and surprise.
2. We detected three kinds of comprehension patterns for the range task:
 - a. No understanding whatsoever--these respondents gave a single number, or no number at all. When probed, their comments indicated that they did not grasp the concept of using a range to reflect their uncertainty.
 - b. A partial understanding they knew that they were expected to provide two numbers, but the numbers referred to something else such as each of the salaries that formed the total.
 - c. A misunderstanding that resulted in reporting income amounts that might have been earned IF PAST CIRCUMSTANCES HAD BEEN DIFFERENT. For example, the highest their income would have been if the spouse had not lost his job.
3. Most respondents had to work hard to recall their income. They needed to construct an answer rather than recalling an already learned answer. This is what we intended. We felt that these conditions would help them understand the concept of response uncertainty. Respondents used a variety of reconstruction strategies,

especially for the five-year change questions.

As in the tests with college students, some of our respondents couldn't or wouldn't follow the task instructions. They gave standard question responses that were outside the high/low range.

Verification of Comprehensibility - As a check on the respondent's final understanding of the uncertainty range concept, at the end of each cognitive interview we asked:

Finally, I'm going to ask you some questions about the amount of paper money (not coins) that you have in your purse (wallet or pockets).

What is the lowest dollar amount of paper money you think you have in your purse (wallet or pockets) at this time?

What is the highest dollar amount you think you have?

And how much paper money do you actually have in your purse (wallet or pockets) at this time?

All respondents answered correctly, in that they gave range and standard answers that met our criteria. And, when we asked them to count their actual paper money, the amount was usually within the range they reported.

The implications of Study One were that we should shorten the questions and do a better job of teaching the range concept.

3.2 Second Study - We conducted a second study in the cognitive laboratory to learn if we could simplify the questions and clarify the task instructions.

3.2.1 Methods - We recruited ten new respondents with characteristics similar to those included in Study One.

Our new strategy was to use two questions to elicit the range boundaries. In addition, we counterbalanced the order of asking for the highest and lowest range values. Furthermore, we counterbalanced asking the standard question before or after the range questions.

An example of a simplified range question is,

What is the highest dollar amount you think this could have been?

3.2.2. Results:

1. Respondents seemed to comprehend the task better but some (albeit fewer) continued to give us answers to the main questions that were either on or outside the high/low interval boundaries.
2. Some respondents still did not understand the range construction task.
3. Some respondents resented the task when we asked for the highest estimate before asking the lowest estimate. None complained when we asked lowest, then highest.

Based on the results from Study Two, we concluded that we still needed to teach the uncertainty range concept more effectively. We needed to retain the short questions and we needed to adopt a consistent order of asking the range boundary items, lowest boundary first, then highest boundary.

3.3 Third Study

3.3.1. Methods - Although we clearly needed more development and testing, our resources were pretty depleted at this point. And we had scheduled the telephone survey for the near future. So we made some final design changes and tested them over the phone on our friends and colleagues.

We introduced a training example for the uncertainty range concept at the beginning of the interview and did not continue until the respondent had correctly reported a standard answer and the endpoints of an uncertainty range that contained the standard answer.

We changed the wording of the standard question to now ask for the best estimate, to further reinforce the idea that the answer could be considered uncertain.

We prompted the interviewer to use specific probes if the respondent's standard answer was outside the uncertainty range, attempting either to extend the range or move the standard answer inside the range.

We wanted to see if the range construction task would go any more smoothly if we asked respondents to report their total household income for 1997 instead of just their wage and salary income. Total income consists of several sources and kinds of income, some of which are difficult to recall exactly. Thus, we hypothesized that a total income first question might make it easier to grasp the uncertainty range concept right away.

We also wanted to examine whether range answers improved if the questions contained some cues about the kinds of income in each category we asked about, for example, regular pay, overtime pay, commissions, bonuses, and tips. Perhaps if we reminded respondents of the many components of earnings and alerted them to the possibility that they may have omitted some and misestimated others, they would be willing to work harder at constructing the uncertainty ranges.

In selected places, we added a question about how confident the respondent was about his best estimate, as a way of introducing the intent of the high/low interval questions that immediately followed.

3.3.2. Results - Asking about total income from all sources instead of total wage and salary income actually made things harder for some respondents and seemed to impede their learning of the range concept. So we dropped that idea.

Respondents seemed to benefit from the extra cues or reminders about what kinds of income to include, even though this material added to the length of the questions. So we kept the extra reminder information in the questions.

The probes we used if the best estimate was outside of the high/low interval worked beautifully, so we kept that procedure.

All respondents did an adequate job with the training example so we kept it at the beginning of the questionnaire.

All respondents readily understood and answered the confidence scale question. However, this would yield a judgment value in the 1-10 range, which cannot be used in the contemplated Bayesian estimate. We could also ask the range questions but, if we follow the established paradigm, we would want to ask the best estimate question before we asked the range questions, which was opposite to what the laboratory study suggested was optimal.

So, we decided to introduce a split-panel experiment into the telephone survey that contrasted two variations on the measurement procedures: The main version (75 percent of the cases) would ask the range questions first, followed by the standard question. The other version would ask the standard question first, then the confidence rating, followed by the two range questions.

4. TELEPHONE SURVEY

The goal of the telephone survey was to obtain a best estimate report of an income amount and a report of the uncertainty range surrounding the estimated amount for several income items. These data will be used in later research to develop improved estimation procedures.

4.1. Sample - With the help of the Census Bureau's Administrative Records Research Staff, we developed a frame of households from commercial and administrative records containing households who filed joint tax returns having wage and salary income for the last five consecutive years. The frame covered the 4 states in which the American Community Survey (ACS) held its first pilot tests. Households interviewed in the ACS tests or for which we could not

obtain current phone numbers were eliminated from the frame. A sample of about 2000 households was drawn from this frame, and each was assigned to an experimental interviewing treatment.

We gave the 2000 names and telephone numbers to our Hagerstown Telephone Facility and asked them to obtain a quota of 500 completed interviews, eliminating households that had become ineligible through retirement, death, divorce or other circumstances that precluded observing the joint wage and salary income on the tax return.

Prior to starting the telephone interviewing, we mailed an advance letter to all 2000 households explaining the survey. For letters returned to us as undeliverable, we notified the Telephone Facility and they removed the household from the sample frame.

4.2 Methods - We used two versions of the questionnaire. Each version asked about wage and salary income and about interest and dividend income for three time periods: the most recent calendar year, 1997, last year, 1996, and the amount of income changes over the last five years (1993-1997). Both versions included questions about characteristics that might correlate with income reporting accuracy, such as:

Who pays the bills? Who fills out the federal tax form?, level of education and age.

Version One of the questionnaire, administered to 75 percent of the eligible, completed cases, asked for the low range boundary first, then the high range boundary, then the best estimate.

Version Two administered to 25 percent of the eligible, completed cases, asked first for the best estimate of the income amount, then the confidence rating, then the lowest range estimate, and finally, the highest range estimate.

Both questionnaire versions began with several questions to help evoke a mental model that included the concept of a best estimate and the range of uncertainty around it. First we established the overall context of income questioning:

We're interested in how people are getting along financially these days. Would you say that you and your family are better off or worse off, financially, than you were a year ago?

Next, we introduced the idea that answers could be uncertain:

We realize people can't report income amounts exactly. So we've designed this survey to make it easier for you. I'll ask you to give me your best estimate. And I'll ask you to report how close your estimate is to the actual value.

Then we used a training question and employed probes, as necessary, to elicit proper answers.

This was the approach for Version One:

To show you what I mean, let's start with a practice question:

What is your best estimate of the average annual income for a family of four in the United States?

What is the lowest the correct value could be?

(If answer is Don't know, ask, Could it be as low as \$1,000? and What is the lowest the correct value could be?)

What is the highest the correct value could be?

(If answer is Don't know, ask, Could it be as high as \$100,000? and What is the highest the correct value could be?)

If the high/low range did not include the best estimate, the interviewer was instructed to use a set of probe questions to bring the discrepancy to the respondent's attention and to provide an opportunity to resolve it. The questions asked depended on the nature of the discrepancy. And then we provided feedback about the successful completion of the task:

Good! You get the idea. Your best estimate is _____. But you feel the correct value could be as low as _____ and

as high as _____. Is this right?

OK, this is how the rest of the questions will go. I'll ask you for your lowest and highest estimates first. Then I'll ask you for your best estimate.

We used a similar approach for Version Two, but asking for the best estimate first, then the confidence rating, then the low boundary and the high boundary value. The feedback followed the Version One questioning pattern.

Telephone interviewing was conducted in May and June of 1998. We held a half-day training session for the telephone interviewers, covering the procedures and concepts, and providing detailed income definition information in case respondents asked about special circumstances.

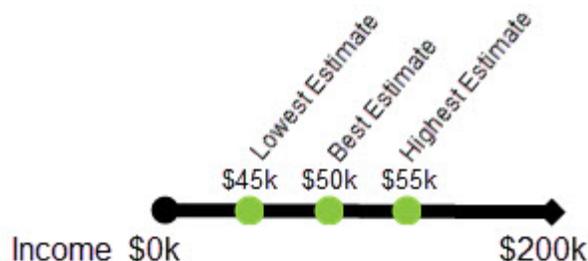
Since the frame information also included data from administrative records about household income, we eventually linked the survey responses to the administrative records to evaluate the validity of the telephone survey responses. As of this writing, we do not yet have the 1997 records information, so we have omitted analyzing both the 1997 and five year change data (that also involve 1997 data). We concentrate on results from the questions dealing with 1996 income.

4.3. Results

4.3.1 Interviewer Debriefing - Our first results come from the interviewer debriefing session. None of the interviewers liked working on this survey. Their comments focused on both their own and the respondents' difficulties in understanding the questions and range concept.

They said they had to repeatedly explain the range concept because respondents often just did not comprehend it. Interviewers had to repeat several questions again and again, as respondents tried to grasp what was being asked. Even though the average interview lasted about 15 minutes, interviewers felt it was too long and too difficult.

4.3.2 Did the telephone survey questions work? - Although our interviewer debriefing suggested that the questions did not work well, the actual data suggest that the interviewers and procedures largely overcame the problems. Recall that our early cognitive tests were plagued by respondents not giving answers, not reporting full ranges, or putting the best estimate outside the high/low range. Our goal was for respondents to specify the range they were sure their income fell within and to report a best estimate within that range.



In the example, this idealized respondent told us that his income could have been as low as \$45,000 or as high as \$55,000 and that his best estimate was \$50,000. The range boundaries are the \$45,000 and \$55,000 values. The best estimate is \$50,000 and it is inside the range.

The telephone survey obtained interviews with 505 households. We now ask how well our procedures worked.

Table 1. How Well Did the Procedures Work? (Percents)

Where is the Best Estimate?	1996 Wage and Salary	1996 Interest and Dividends
INSIDE the range	72	57
On the range BORDER	21	22
OUTSIDE the range	3	2

MISSING or No Range	4	19
Total Percent (n=505)	100	100

The 1996 income response data suggest we are well on our way to evolving a workable set of procedures (Table 1). For both kinds of 1996 income, Wages and Salaries and Interest and Dividends, over 3/4 of the respondents gave answers that conformed to the intended format, either the best estimate was inside the range or equal to one of the extreme values (on the border).

Notice that these procedures managed to keep the best estimates from going outside the range, undoubtedly due to the computer assisted probe questions that were automatically displayed when an out-of-range problem occurred.

The single disappointment is the high rate of missing data for the interest and dividends item, almost 20 percent. These probably result from metacognitive judgments of not knowing, followed by an unwillingness to try further recall. Clearly we have some additional work to do to persuade respondents to keep trying to recall interest and dividend information and to complete that particular kind of reporting task. Perhaps furnishing additional cues about the likely sources of dividend and interest income would help.

4.3.3. Did subgroups have trouble with the procedures? - For the remaining analyses, we examine whether particular subgroups experienced special difficulties with the procedures. We will look at correlations of evaluation variables with five group characteristics:

Status Groups

- Whether the respondent pays the bills or not;
- Whether the respondent does the annual federal tax forms.

Procedural Groups

- Version 1 or Version 2 of the questionnaire.

Demographic Groups

- Respondent age; Respondent education level.

Table 2. Correlations of Group Characteristics with Conforming Responses
(Conforming = Best estimate is inside or on the range border)

Group	1996 Wage & Salaries (n = 505)	1996 Interest & Dividends (n = 505)
R pays the bills	+.00	-.09
R does the taxes	-.00	+.04
Questionnaire Version	+.04	-.04
Age	-.04	-.16*
Education	+.04	+.07

Table 2 shows that almost none of these characteristics correlated with conforming to our procedures, which we defined as: giving a range and a best estimate inside the range or on the border of the range. The data suggest that older respondents may have had a little more trouble meeting expectations for the dividend and interest question.

4.3.4. Were the best estimates accurate? - We define accuracy in terms of how close the survey response is to the 1996

entry on the family's federal income tax form. If respondents asked for definitions during the survey, we gave them the definitions of income components (what to include and exclude) that were consistent with federal personal income tax definitions.

Table 3. Correlations of Survey Best Estimate and Tax Form Income Amounts

Survey and Tax Form:	Correlation (R)	R-squared
Wage & Salary (n = 490)	.68*	.46
Interest & Dividends (n = 408)	.77*	.59

Table 3 shows the correlations between the survey best estimates and tax form responses for the two kinds of 1996 income. The (untransformed) survey responses do correlate moderately well with the tax form values, even after more than a year had elapsed. Ideally the responses would account for 100% of the variance in the tax form values; the R-squared values in the table suggest that these responses account for 50-60 percent, which is not bad, but far short of what some data users assume surveys achieve.

Are some subgroups of respondents more accurate than others? We obtained the subgroup correlations with an ERROR variable that we defined as:

$$\text{ERROR} = (\text{Survey Value} - \text{Tax Form Value}) / \text{Tax Form Value}$$

The numerator reflects the discrepancy between the survey and tax values. The absolute value operator makes it possible to consider both positive and negative deviations to be errors. The denominator acts to standardize the discrepancy values so that especially high or low incomes don't distort the score relative to other people's scores: Note that the largest error score due to completely underreporting income is 1. For symmetry and to control the effects of outliers on correlations, we arbitrarily set the highest error score for income overreporters to be 1 also. Tax form income values of zero excluded the case from receiving an error score.

Table 4. Correlations with Best Estimate Error

Group	1996 Wage & Salaries (n = 455)	1996 Interest & Dividends (n = 374)
R pays the bills	-.04	-.02
R does the taxes	-.07	.00
Questionnaire Version	-.04	+.06
Age	+.22*	-.23*
Education	-.02	-.06

Table 4 shows that error scores are not correlated with most of our subgroup variables, There is a strange pattern of findings for age: older respondents seem to make larger errors on the wage and salary variable, and smaller errors on the interest and dividends variable.

4.3.5 Do the reported ranges contain the criterion values? Table 5 shows that between 66-71 percent of the reported 1996 ranges included the tax value, a respectable showing. So the ranges do contain accurate information that should be useful in improving the population estimates of income amounts.

To construct the score, we assigned a range-accuracy value of 1 if the tax value was inside the reported range or equal to one of the border values. Otherwise, if the range was reported, the score was zero. We ignored the cases where the respondent did not provide a range. If they were included as incorrect, the percent correct values would be somewhat

smaller for wage and salary income and considerably smaller for interest and dividend income.

Table 5. Do the Survey-Reported Ranges Include the Tax Form Value?

	1996 Wage and Salary (N=484)	1996 Interest and Dividends (N = 407)
Percent of Ranges that include the tax form value	66%	71%

For subgroups, age is again negatively related to accuracy for both variables (Table 6). There is no effect of the respondent's financial role, questionnaire version or education.

Table 6. Does The Range Include True Value?
(Correlations)

Group	1996 Wage & Salaries (n = 484)	1996 Interest & Dividends (n = 407)
R pays the bills	+.01	-.02
R does the taxes	-.00	+.01
Questionnaire Version	-.04	-.03
Age	-.12*	-.13*
Education	+.06	+.02

5. DISCUSSION AND CONCLUSIONS

The computer-assisted telephone survey results suggest that it is possible to construct questioning procedures that result in respondents reporting a confidence range and best estimate for components of their household income.

However, cognitive laboratory research suggests that such concepts are difficult for respondents to grasp. Indeed, we had to make some very drastic changes in procedures to come as far as we have.

For example, to get workable questions, we had to break up long paragraphs into short, succinct items. In order to reduce the short-term memory load, we had to ask for one answer at a time rather than all 3 parts at once. And to remove the test-like quality, we changed the form of speech from imperative instructions to actual questions.

Even at this stage, telephone interviewers report that they must repeat the questions more than once and they must use the special probes often to rectify problem situations. If interviewers dislike a survey, they may not do as good a job as otherwise. So, there is still room for procedural improvement.

Analyses suggest that there is also room for improving the accuracy of the answers. Correlations with criterion values were moderate although, encouragingly, the ranges appeared to include the criterion value more than 2/3rds of the time.

There is valid information in these answers, so it seems worthwhile to proceed with this research as planned. Our next step will be to evolve Bayesian estimators of the population mean that are more accurate and more precise than ordinary estimators.

Indeed, it is probably true that no procedures can be devised that enable perfect reporting by respondents. So the major technical improvements may come from innovative estimation methods (or from the direct use of information in administrative records).

If we do attempt new research with questionnaires, we might try two different approaches:

One approach might involve capturing different forms of the metacognitive information, possibly involving confidence judgments instead of range definitions, if appropriate ways of using such information in estimates can be derived.

A second approach might involve trying to change the metacognitive judgments that lead respondents not to try to recall difficult information such as dividend and interest income. Providing questions that contain examples (recognition cues) of the kinds of income to be included, may produce more positive metacognitive judgments (feelings of knowing) and result in more recall effort.

References

Koriat, Asher (1994), Memory's Knowledge of Its Own Knowledge: The Accessibility Account of the Feeling of Knowing. In Metcalf, Janet and Arthur P. Shimamura (eds.) *Metacognition: Knowing About Knowing*. Cambridge: The MIT Press.

Metcalf, Janet, B. L. Schwartz and S. G. Joaquim (1993), The Cue-Familiarity Heuristic in Metacognition, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 19, pp. 851-861.

Metcalf, Janet and Arthur P. Shimamura (eds.) (1994), *Metacognition: Knowing About Knowing*. Cambridge: The MIT Press.

Reder, Lynne and C. D. Schunn (1996), A Metacognition Does Not Imply Awareness: Strategy Choice is Governed by Implicit Learning and Memory, in L. M. Reder (ed.) *Implicit Memory and Metacognition*, Mahwah NJ: Lawrence Erlbaum Associates.

Sporer, S. L., S. Penrod, D. Read and B. Cutler, Choosing, Confidence, and Accuracy: a Meta-analysis of the Confidence-Accuracy Relation in Eyewitness Identification Studies. *Psychological Bulletin*, 1995, vol. 118, pp. 315-327.