

**Pre-testing Sensitive Questions:  
Perceived Sensitivity, Comprehension,  
and Order Effects of Questions about  
Income and Weight** August 2005

Margaret K. Vernon  
Bureau of Labor Statistics, Room 1950,  
2 Massachusetts Ave, NE  
Washington, DC 20005  
[vernon\\_m@bls.gov](mailto:vernon_m@bls.gov)

**Keywords:** Sensitive questions,  
cognitive testing, survey development

**Abstract**

Several new questions dealing with eating habits, family income, eligibility for food stamps, and body mass index are being considered as a supplement to the American Time Use Survey. This paper discusses methods used for pre-testing sensitive questions and addresses concerns about possible order effects of the sensitive questions. New questions were pre-tested with two phases of cognitive interviews designed to assess comprehension, perceived sensitivity of questions, and order effects. Special attention was paid to the questions about income and weight as they were anticipated to be the most sensitive questions proposed for the supplement. Cognitive testing results suggested that the income question was difficult for participants to respond to and was perceived to be the most sensitive question in the supplement, followed by weight. The second phase of testing randomly varied the order of income and weight questions to assess order effects in perceived sensitivity. Income was always reported to be more sensitive than weight, whether it came before or after weight. Yet, the weight question generated item nonresponse and more visual signs of discomfort from

participants. Cognitive testing methods for new item development and results of testing are discussed.

**Introduction**

The American Time Use Survey (ATUS), which began production in 2003, is a household survey that uses a computer assisted telephone interview as the mode of data collection (see Horrigan & Herz, 2004 for a more detailed description). The interview gathers data from one designated person in a selected household about his or her time spent in daily activities during the previous 24- hour period. Open-ended descriptions of activities, duration (or actual starting and stopping times), and questions about with whom and where the activity occurred are also collected in the form of a time diary. In addition, several summary questions immediately following the diary ask about work, childcare, volunteer activities and trips away from home for two or more nights in a row. Recently, several new questions that are designed to measure eating habits, weight, and income have been proposed to be added to the interview in the form of a supplement. These questions were designed to investigate the relationship between these variables and how people use their time – for example, the possible relationship between income and weight and time spent in food preparation and shopping.

A total of 11 questions (or possibly less, depending on skip patterns) have been proposed for the supplement. Before these questions could be added to the ATUS, the questions needed to be pre-tested in order to assess participants' ability to comprehend and respond to them. Cognitive interviewing, a

qualitative technique for assessing question quality, was chosen as the primary method for pre-testing these questions. Cognitive interviewing is a technique that studies how participants understand, mentally process, and respond to questions being asked in the survey (Willis, 2005). The goals of testing the supplement questions were to find out more about whether or not participants comprehend the questions as they were intended and how easy or difficult it was to process and respond to the questions asked (i.e., is it possible for people to respond in a consistent and coherent way that will produce quality data). In addition to these goals, the cognitive testing of the supplement questions had a third goal. Specifically, two of the questions proposed in the supplement were identified as potentially sensitive in nature, and the cognitive testing was designed to assess whether these questions would cause an adverse reaction in participants and potentially generate item nonresponse.

The two questions that were identified as potentially sensitive were questions about weight and income. Previous research has shown that questions about income are among the most sensitive questions that can be asked in a survey. Laumann, Gagnon, Michael, & Michaels (1994) found that income questions were perceived to be even more personal and sensitive than some questions about sexual behavior. Therefore, it was predicted that the proposed income question might be seen as very sensitive and cause participants discomfort. In addition, a question about how much a respondent weighs was predicted to generate similar reactions.

In addition to weight and income potentially being perceived as sensitive questions, it was anticipated that comprehension, and, thus, ability to respond accurately to the income question might be difficult. The income question initially proposed for the supplement was a highly detailed question that included a list of items to keep in mind and gave a time frame (see Methods Phase 1 section for exact question wording). Long lists of items to include or not include are difficult for participants to keep in mind when answering a survey question (Fowler, 1995). In addition, the specific income questions that government agencies typically prefer have been shown to have comprehension and data quality problems (Moore, Stinson, & Welniak, 2000). Therefore, it was anticipated that the wording of the income question would be difficult to comprehend, and that the question would be difficult to answer correctly.

This paper focuses on the cognitive testing results of the two questions that were identified as potentially sensitive: weight and income. Two phases of testing were undertaken that assessed comprehension, response quality, and perceived sensitivity of the weight and income questions. The first phase of testing took a primarily qualitative approach. In the second phase of testing, a more quantitative approach was taken in order to assess whether respondent gender or the order of the sensitive questions were related to the perceived sensitivity of the questions. Results from both phases will be discussed.

### **Method Phase 1**

The first phase of cognitive testing included 8 participants who were recruited from a database of potential participants kept by BLS for recruiting purposes. Most of the participants had either responded to an advertisement in the newspaper that was designed to recruit participants for cognitive studies or were friends or relatives of BLS employees. All 8 participants lived in the greater Washington, DC area. Five of the participants were female, three were male. Six of the participants were African American, one was Hispanic, and one was White/Caucasian. Participants ranged in age from 23 to 64, and their household income levels ranged from around \$6,000 to \$75,000 per year. All participants traveled to the BLS cognitive lab to participate in the study, and all were paid \$35 for their time.

When participants arrived at the BLS cognitive lab, they were escorted to a room with a small round table. The purpose of the study was explained and participants were informed of the voluntary and confidential nature of the study. Since the ATUS is a telephone interview, the first portion of the interview was conducted over the phone. To make the experience of answering the supplement questions as similar to an actual ATUS interview as possible, the cognitive lab room was equipped with a phone, and the interviewer went into another room and called the participant on the phone. The interviewer then administered the survey, which included the actual ATUS time diary and the proposed supplement questions. The time diary portion of the survey, which is considered the core of the ATUS interview, was administered first. Then, the interviewer administered the

supplement. When fielded, the supplement questions will come at the end, so the telephone interview experience mimicked the experience that participants will have when they are contacted for an interview.

After the phone portion of the cognitive interview, the interviewer returned to the room and engaged the participant in a retrospective verbal probing interview (see Willis, 2005, for a review of cognitive interviewing techniques). The two questions that are the focus of this discussion, the questions about weight and income, were embedded in the supplement. Thus, several questions were asked before the weight and income questions and introductions were provided for both. In the first phase of testing, the weight question was asked second to last and the income question was asked as the very last question in the interview.

In the supplement, introductions to the weight and income questions both alert the participant to what is coming next, and provide a rationale for asking these questions. Research has shown that the likelihood of answering sensitive questions is related to the extent to which the participant can see a relation between the question and the objectives of the research (Fowler, 1995). The sequence of questions leading up to the weight and income questions was as follows:

- Alright, we're almost done here. I'm going to switch topics and ask you a few final questions about your physical health that might affect how you use your time.

- In general, would you say that your physical health is very good, good, fair, poor, or very poor?
- How tall are you without shoes?
- How much do you weigh without shoes?
- The final question is about your household income.
- I'd like to ask you about the total combined income of your family during the last 12 months. This includes money from jobs, net income from business, farm or rent, pensions, dividends, interest, social security payments and any other money income received (by members of this FAMILY who are 15 years of age or older.) What was your total household income in the last 12 months?

Several specific probes were developed to examine the comprehension, response type and quality, and sensitivity of the weight and income questions. The interviewer, during the retrospective probing portion of the interview, said something like, "Alright, then I asked you about your weight. I asked How much do you weigh without shoes?" The interviewer would confirm that the participant recalled being asked this question and then would begin to probe using a script developed ahead of time. In addition, if there had been a cue that the participant was uncertain, paused, or displayed some other behavioral cue, the interviewer would spontaneously probe about this. The standard probes were asked as follows:

### **Weight**

- Can you tell me how you arrived at your answer for your weight without shoes?

- When was the last time your weight was measured?
- Do you know your weight **exactly** without shoes?
  - Would you say your answer was exact or within a couple of pounds?
- How did you react to being asked this question?
- Does asking about health, height and weight together make sense to you?
- If you had to give a good reason for asking these questions, what would it be?

### **Income**

- How did you arrive at your answer for this question?
  - What time period did you think about?
  - Was this amount before or after tax deductions?
- Which people did you include when estimating your family's income?
- What sources of income did you count?
- How accurate an estimate of your household income were you able to give me?
  - Would you say it was very accurate, accurate, or not very accurate?

How do you think people will react to being asked about their income?

Once all of the supplement questions had been probed about, the interviewer thanked participants for their time, asked if they had any further questions, and concluded the session.

### **Results Phase 1** **Comprehension**

The question about weight was well understood. No response problems came

up in the retrospective probing that suggested that the weight question was interpreted in any other way than what was intended.

The income question was less well understood. Several comprehension problems occurred. The question was intended to generate an estimate for income before taxes. Seven of eight people gave a response for their income before taxes but one person responded with household income after taxes. In addition, there was confusion about the time frame for when income should be reported. Four out of eight respondents said that they reported their household income for the last calendar year. Four said that they reported income for the last 12 months, as was intended by the question.

### **Response Quality**

The response quality and validity of the responses to weight and income are unknown. In order to assess how accurate the weight reports were, an actual weighing in would have been necessary. This was seen as excessive for the purposes of this cognitive testing study. In addition, no record checks were done to assess the validity of the responses obtained for income. However, based on the variability of the responses given, with some people reporting income before taxes and some reporting after taxes, in addition to reporting for different time frames, it can be assumed that there will be variability in the data.

### **Sensitivity**

Although sensitivity was not quantitatively assessed during the first phase of testing, five out of eight people expressed that the weight question was

“very sensitive”. One person spontaneously said that weight is more sensitive than income (without being asked to compare the two). One person refused to answer the question. Almost all participants provided some sort of behavioral cue that indicated they felt uncomfortable with the question. For example, there were pauses, nervous laughter, or explanations such as, “well, my doctor did tell me that I need to lose some weight”.

Surprisingly, there did not seem to be as much discomfort with the income question as was generated by the weight question. No one refused to answer the income question. Three out of eight people said that income could be a little sensitive. There were no overt behavioral cues that suggested discomfort.

As was discussed in the introduction, income tends to be perceived as one of the most sensitive and personal questions that can be asked about in a survey (Laumann, Gagnon, Michael, & Michaels, 1994). Therefore, it was surprising that the weight question generated a refusal while no one refused to answer the income question. In addition, weight seemed to generate stronger verbal and behavioral statements of discomfort. One explanation for why weight could have been perceived as more sensitive is that it came before income. A hypothesis, therefore, is that the first sensitive question asked would be perceived as the most sensitive and the second would generate less discomfort because the “shock” of the first question would desensitize participants for later sensitive questions. Phase 2 of testing addressed this hypothesis with an experimental

design and a more quantitative approach to measuring sensitivity and discomfort to the weight and income questions.

### **Method Phase 2**

A total of 16 participants (8 male, 8 female) were recruited for Phase 2 of the cognitive study. Participants were again recruited from the BLS participant database or were family or friends of BLS employees. Ten of 16 participants were African American and six were White/Caucasian. The participants ranged in age from 21 to 62 years of age. Household income of participants ranged from \$6,468 to \$210,000 per year.

The second phase of testing employed an experimental design. The 16 participants were randomly assigned to two groups. Assignment, however, was not completely random because gender was balanced in the two groups. Therefore, each group had four females and four males.

The procedures for the second phase of testing were similar to procedures for the first. Participants were first given the telephone interview (time diary first, followed by the supplement questions). However, in Phase 2 the order of the supplement questions varied. Participants in Group 1 received the supplement questions in the same order as they were received in the first phase of testing. That is, participants answered the weight question first, followed by the income question. Group 2 received the weight and income questions in the opposite order. For Group 2 the income question came first, followed by the health introduction, general health question, height, and then weight (see Methods Phase 1 for exact question wording).

Immediately following the telephone portion of the interview, participants were given a paper and pencil survey designed to assess perceptions of sensitivity about the questions in the supplement. Several questions in the supplement (not just weight and income) were asked, and all questions appeared on the paper survey in the same order as they were asked about in the phone interview. Other questions asked about in the survey covered topics such as secondary eating and drinking, food preparation, and family participation in the food stamp benefit program.

In the paper survey, two questions were asked about reactions to supplement questions. The question that was asked in the supplement portion of the telephone interview was restated. Then participants were asked, "When this question was asked, how did you feel"? Response options ranged from 1 (Very Uncomfortable) to 5 (Very Comfortable). The second follow-up question was, "How personal do you think the information is that this question is requesting"? Response options ranged from 1 (Not at all personal) to 5 (Very personal).

After the paper and pencil survey, the same retrospective probing technique was undertaken. Since question wording did not change from Phase 1 to Phase 2, the same probes used in Phase 1 were used in Phase 2.

## **Results and Discussion**

### **Comprehension**

Similar results as reported in Phase 1 were found in Phase 2. Weight was well understood. Several more idiosyncrasies

were found with how the income question was interpreted and answered. In addition to people reporting income before and after taxes and using different time frames, one person combined before-tax income with other after-tax sources. Two people had irregular incomes and had difficulty answering the question. One person forgot to include dividends (even though this was cued). One person gave a monthly income instead of a 12-month income, and two people gave ranges that were so large the response was meaningless.

### **Response Quality**

Although there is no way to know for certain about the validity of the weight responses, the variability in how people answered the income question raised concerns about a large amount of potential error in the income data. Major issues to consider when designing an income question include carefully considering whether the desired estimate is before or after taxes. In addition, the question time frame needs to be clear. In this case, even though the question clearly stated, "last 12 months," many people gave an estimate for the last calendar year (possibly because this is the way we are trained to think about yearly income for tax purposes). In addition, long lists that contain many items, especially if read over the phone, might be difficult for people to remember. This was evidenced by one person forgetting to include dividends, even though instructions to include dividends were clearly stated. There is some research that suggests that when long lists are presented in an auditory mode, people tend to forget items on the list, especially if they are closer to the beginning of the list (Krosnick & Alwin,

1987). Finally, there are many people who have incomes that are not stable from month to month, and answering questions about household income over the last 12 months might prove a very difficult task for these individuals. Paying careful attention to the reference period might help these types of individuals to better respond to the question.

### **Sensitivity**

Similar results were found in Phase 2 about the sensitive nature of weight and income. As in Phase 1, in Phase 2 one person refused to answer the weight question but no one refused income. Interestingly, there were no significant differences in how comfortable people felt when asked about weight compared to income ( $t = .72$ ,  $p = .48$ ). However, people did feel that income was significantly more personal than weight ( $t = -2.22$ ;  $p = .04$ ). Even though one person refused to answer weight in each phase of testing, and no one refused income, weight and income were reported to generate similar levels of discomfort and income was perceived to be more personal than weight.

ANOVA's were employed to test possible order effects of the two sensitive questions. Specifically, it was predicted that the first sensitive question received would be perceived as the more sensitive question. Therefore, it was hypothesized that Group 1 would feel that weight was more sensitive compared with Group 2, and that Group 2 would feel that income was more sensitive than Group 1. Results showed that there were no differences between the two groups in how comfortable or

personal weight and income were perceived to be (**weight:** comfort:  $F = 1.41$ ;  $p = .26$ ; personal:  $F = 0.17$ ;  $p = .69$ ; **income:** comfort:  $F = 0.29$ ;  $p = .56$ ; personal:  $F = 0.00$ ;  $p = .98$ ). In fact, even though the differences were not significant, both groups felt slightly less comfortable with reporting income compared to weight and both felt that income was slightly more personal, regardless of the order questions were asked (see Tables 1 and 2 for means and standard deviations).

Finally, it was predicted that women would feel that weight was more sensitive than men. ANOVA results suggested that there were no differences between men and women in how comfortable they felt when asked the weight question ( $F = 2.50$ ;  $p = .14$ ). In addition, there were no significant differences in how personal the weight question was perceived to be ( $F = 1.40$ ;  $p = .26$ ). Mean differences suggest that the trend is for women to feel less comfortable, and to feel that weight is more personal, but the differences were not significant (see Table 3 for means and standard deviations). This could be due to the small sample sizes of men and women reporting.

These results are in line with previous research that has found that income is one of the most sensitive questions that can be asked in a survey (Laumann, Gagnon, Michael, & Michaels, 1994). Most differences were not significant, perhaps due to the small sample sizes, but there were small mean differences in levels of comfort and how personal the questions were perceived to be in favor of weight. That is, participants felt less comfortable with income than weight

and felt that income was more personal than weight. These results were surprising given the fact that two people refused to answer the weight question while no one refused income. In addition, there were more behavioral cues of discomfort to the weight question than to income. One potential explanation for these results could be that although individuals in this study felt personally uncomfortable with talking about their own weight, there is a broader social norm surrounding the inappropriateness of talking about income in public. Therefore, individuals could have felt personally uncomfortable talking about their weight but feel that talking about income, in general, is more of a personal, inappropriate, or taboo topic. The fact that income and weight seemed to generate similar levels of discomfort, while income was said to be more personal than weight, seems to back up this hypothesis. However, the finding that more people refused weight, a true behavioral indicator, could indicate that weight is the more sensitive of the two items. On the other hand, the finding that more people refused weight compared with income could also have been an artifact of having a small, non-representative sample. This could have been a chance finding. If the sample had been larger, perhaps more people would have refused income than weight.

### **Future Directions**

The income question tested here will not be fielded in the American Time Use Survey due to the difficulties that arose with comprehension and response to the question during testing. An alternative income question that the Economic Research Service suggested, and that was tested by BLS, will be fielded (see

Appendix A for alternative income question wording). When the questions about weight and income are added to the survey, this will provide nationally representative data with which to test whether there is more item nonresponse to the weight question or the income question.

## Reference

- Fowler, F. J. Jr. (1995). Improving survey questions: Design and evaluation. Thousand Oaks, CA: Sage Publications.
- Horrigan, M. & Herz, D. (2004). Planning, designing, and executing the BLS American Time Use Survey. Monthly Labor Review, 127 (10), 3-19.
- Krosnick, J. & Alwin, D. (1987). An evaluation of a cognitive theory of response order effects in survey measurement. Public Opinion Quarterly (51), 201-219.
- Laumann, E. O., Gagnon, J. H., Michael, R. T., & Michaels, S. (1994). The social organization of sexuality: Sexual practices in the United States. Chicago, IL: University of Chicago Press.
- Moore, J. C., Stinson, L. J., Welniak, E. Jr., (2000). Income Measurement Error in Surveys: A Review. Journal of Official Statistics, 16 (4), 331-362.
- Willis, G. B. (2005). Cognitive interviewing: Tools for improving question design. Thousand Oaks, CA: Sage Publications.

**Appendix A**  
**Alternative Income Question Wording**

**>Question Text**

The next question is about your household income.

**>Question Text**

fill: Last month, was your total household income before taxes more or less than [fill 1] per month?

\*Probe (if necessary): that would be [fill 2] per year.

1. More
2. Less
- 2. Don't know
- 3. Refused

\*If less, go to next income question, if more, skip next income question

**>Fill Instructions**

1. If total number of people living in the HH = 1; fill \$1, 436.00  
If total number of people living in the HH = 2; fill \$1,927.00  
If total number of people living in the HH = 3; fill \$2,416.00  
If total number of people living in the HH = 4; fill \$2,907.00  
If total number of people living in the HH = 5; fill \$3,397.00  
If total number of people living in the HH = 6; fill \$3,888.00  
If total number of people living in the HH = 7; fill \$4,377.00  
If total number of people living in the HH = 8; fill \$4,868.00  
If total number of people living in the HH > 8; fill \$4,868.00 + (\$491.00 X each additional HH member)
2. If total number of people living in the HH = 1; fill \$17,232.00  
If total number of people living in the HH = 2; fill \$23,124.00  
If total number of people living in the HH = 3; fill \$28,992.00  
If total number of people living in the HH = 4; fill \$34, 884.00  
If total number of people living in the HH = 5; fill \$40,764.00  
If total number of people living in the HH = 6; fill \$46,656.00  
If total number of people living in the HH = 7; fill \$52,524.00  
If total number of people living in the HH = 8; fill \$58,416.00  
If total number of people living in the HH > 8; fill \$58,416.00 + [(\$491.00 X 12) (each additional HH member)]

**>Question Text**

fill: Was it more or less than [fill 1] per month?

\*Probe (if necessary): that would be [fill 2] per year.

1. More
2. Less
- 2. Don't know
- 3. Refused

**>Fill Instructions**

1. If total number of people living in the HH = 1; fill \$1,009.00  
If total number of people living in the HH = 2; fill \$1,354.00  
If total number of people living in the HH = 3; fill \$1,698.00  
If total number of people living in the HH = 4; fill \$2,043.00  
If total number of people living in the HH = 5; fill \$2,387.00  
If total number of people living in the HH = 6; fill \$2,732.00  
If total number of people living in the HH = 7; fill \$3,076.00  
If total number of people living in the HH = 8; fill \$3,421.00  
If total number of people living in the HH > 8; fill \$3,421.00 + (\$345.00 X each additional HH member)
2. If total number of people living in the HH = 1; fill \$12,108.00  
If total number of people living in the HH = 2; fill \$16,248.00  
If total number of people living in the HH = 3; fill \$20,376.00  
If total number of people living in the HH = 4; fill \$24,516.00  
If total number of people living in the HH = 5; fill \$28,644.00  
If total number of people living in the HH = 6; fill \$32,784.00  
If total number of people living in the HH = 7; fill \$36,912.00  
If total number of people living in the HH = 8; fill \$41,052.00  
If total number of people living in the HH > 8; fill \$41,052.00 + [(\$345.00 X 12) (each additional HH member)]

**Appendix B**  
**ANOVA Tables**

**Table 1**

Means and standard deviations of weight sensitivity measures by group (see results phase 2 section for ANOVA F test reports)

|         | Comfort                  | Personal    |
|---------|--------------------------|-------------|
| Group 1 | 4.25 (1.39) <sup>a</sup> | 3.00 (1.69) |
| Group 2 | 3.43 (1.27)              | 3.29 (0.76) |

<sup>a</sup> mean (standard deviation)

**Table 2**

Means and standard deviations of income sensitivity measures by group (see results phase 2 section for ANOVA F test reports)

|         | Comfort                  | Personal    |
|---------|--------------------------|-------------|
| Group 1 | 3.75 (1.16) <sup>a</sup> | 3.88 (1.55) |
| Group 2 | 3.43 (1.13)              | 3.87 (0.38) |

<sup>a</sup> mean (standard deviation)

**Table 3**

Means and standard deviations of weight sensitivity measures by gender (see results phase 2 section for ANOVA F test reports)

|         | Comfort                  | Personal    |
|---------|--------------------------|-------------|
| Females | 3.38 (1.60) <sup>a</sup> | 3.50 (1.20) |
| Males   | 4.43 (0.79)              | 2.71 (1.38) |

<sup>a</sup> mean (standard deviation)

