

SUBJECTIVE ASSESSMENTS OF ECONOMIC WELL-BEING AND PRELIMINARY FINDINGS FROM MIAMI

Thesia Garner, Linda Stinson, and Stephanie Shipp, Bureau of Labor Statistics
Thesia Garner, BLS , 2 Massachusetts Ave. NE, Wash, DC 20212

Key words: minimum income question, income evaluation question, delighted/terrible question

Surveys have long been used to measure both objective and subjective phenomena. Objective realities, such as one's date of birth or educational attainment, are matters of public record. Subjective reality, on the other hand, is totally private, it can be reported only by the person experiencing it. In this study we are concerned with respondents' private views of their income. These views may be included in their subjective assessments of income and may help to determine whether or not they experience some sense of economic well-being. Since most American adults are confronted daily with the delicate task of balancing their income and expenses, it seems reasonable to expect them to have a rather well-developed budgetary sense.

Subjective assessments of economic well-being and income, in particular, can play an important role in social analysis. For example, respondents' subjective assessments of the contribution made by their income to the attainment and maintenance of their social roles can be used to produce utility functions. This would be an alternative to trying to plot utility functions based on using only the value of each individual's income, given his or her unique circumstances, a most difficult task.

The major survey problem when measuring subjective phenomena is the susceptibility to nonsampling error. Any change, temporary or permanent, in or around the responding subject (mood, experience, weather) may alter that subject's perceptions, interpretations, and conclusions—if only for a moment. Likewise, the context of the interview itself, the order in which questions are presented, and even the characteristics of the interviewer may affect the respondent's subjective assessment. In such cases, it would be difficult, if not impossible, to determine whether changes in subjective assessments represent true variation resulting from a change within the subject or the intrusion of measurement error into the survey process. As a result, the effective use of subjective questions must absolutely depend upon clear, precise, and unambiguous language, non-arbitrary response categories, and clearly defined concepts.

For the purposes of this study, the clarity and precision of four subjective questions are tested:

Minimum Income Question (MIQ)

Living where you do now and meeting the expenses you consider necessary, what would be the smallest income (before any deductions) you and your family would need to make ends meet?

Minimum Spend Question (MSQ)

In your opinion, how much would you have to spend each month in order to provide the basic necessities for your family?

Income Evaluation Question (IEQ)

Which after-tax monthly income would you, in your circumstances consider to be very bad? bad? insufficient? sufficient? good? very good?

Delighted/Terrible Question (D/T)

Which of the following categories best describes how you feel about your family income (or your own income if you are not living with relatives)? Do you feel delighted, pleased, mostly satisfied, mixed, mostly dissatisfied, unhappy, or terrible?

Once we gain a better understanding of these questions, greater progress can be made to combine subjective assessments and objective measures of economic well-being.

The remainder of this paper is organized into five major sections: previous research and U.S. Government involvement, study design, methodology, preliminary results, and summary and conclusions.

Previous Research and U.S. Government Involvement

During the twenty years in which subjective questions have been used to assess the economic well-being of individuals and households, there have been only a few studies done on the reliability or validity of these particular questions. Most notable are the studies by Andrews and Withey (1976), Kapteyn (1979), Antonides et al. (1986), and Morissette and Poulin (1991).¹ Other researchers have examined these

¹ F. M. Andrews and S. B. Withey. *Social Indicators of Well-Being*. New York: Plenum Press, 1976, p. 343.

questions and have provided insight concerning their meaning.²

During this same time period, the U.S. government moved forward and included subjective assessments of health status and health-related work limitations in federally sponsored surveys. The move to include subjective measures of the impact of income and financial resources in federally sponsored surveys, however, was not as readily endorsed.

The first foray of subjective assessments of income in federally sponsored surveys came when both the MIQ and the D/T question were included in the Bureau of the Census' 1979 Research Panel of the Income Survey Development Program.³ The MIQ was also added to the 1982 Consumer Expenditure Interview Survey⁴ based upon the recommendation of the Expert Committee for the Bureau of Labor Statistics' (BLS) Family Budget Revisions.⁵ Consequently, the committee recommended a major effort to evaluate and perfect a survey methodology that would permit a paradigm shift away from the notion that official experts can (and should) define what the populace needs in order to get along or prosper. The new measurement task would be directed instead toward finding stable, reproducible estimates of what ordinary people experience with their family budgets (p. 9). Proposals by the committee and subsequent groups met with OMB resistance. Later, OMB became more

receptive to these questions but before they could be implemented, they recommended laboratory and field testing. Further supporting these efforts, the use of subjective measures was endorsed in the Committee on National Statistics report *Measuring Poverty, A New Approach*⁶.

In the Spring of 1996, BLS management determined that this first stage of cognitive testing might proceed, with the understanding, that upon completion, there would be a review of the results and decisions made concerning further implementation. The format of the test and the questions to be asked were developed and pre-tested in the BLS cognitive lab. Staff from the BLS and the Michigan Survey Research Center, under contract with BLS, are collecting the data from five areas in the U.S.

At about the same time, OMB agreed that the D/T question, the MIQ, and the MSQ would be collected in a SIPP module^{7,8}. These data were collected by the Bureau of the Census.

In an effort to coordinate activities and to meet the OMB request, the BLS and Census Bureau are jointly funding the cognitive testing. Once the SIPP field data are available, BLS will work with the Census Bureau in analyzing those data. Then results from the cognitive tests and the SIPP data will be used to provide an overall assessment of the subjective questions under examination.

Study Design

The study is designed to address four main research questions:

1. How do respondents interpret such terms as "minimum income," "sufficient income," "necessary expenses," and "monthly household income?"
2. Are there potential order effects when asking respondents to make subjective assessments? For example, when respondents are asked to assess satisfaction with their income, are their ratings affected by questions immediately preceding those asked about their expenditures?
3. How do respondents use response categories such as "good/bad,"

² Peter Saunders and George Matheson, *Perceptions of Poverty, Income Adequacy and Living Standards in Australia*. Social Policy Research Centre Reports and Proceedings, No. 99, The University of New South Wales, April 1992.

Thesia I. Garner and Klaas de Vos, "Income Sufficiency vs. Poverty: Results from the United States and the Netherlands," *Journal of Population Economics*, vol. 8, 1995 pp. 117-134. Also see: Klaas de Vos and Thesia I. Garner, "'An Evaluation of Subjective Poverty Definitions: Comparing Results from the U.S. and the Netherlands,'" *Review of Income and Wealth*, Series 37, Number 3, September, 1991.

³ For a review of the "history" of self-assessments in potential federal household surveys, see Denton R. Vaughan, "Self-Assessments of Income Needs and Financial Circumstances: Two Decades of Seeking a Place in Federal Households Surveys," paper presented at the Joint Statistical Meetings, Sections on Government Statistics, Social Statistics, and Survey Research Methodology, Chicago, Illinois, August 5, 1996.

⁴ See deVos and Garner (1991) and Garner and deVos (1995) for analyses of these data.

⁵ Harold W. Watts. "Special Panel Suggests Changes in BLS Family Budget Program," *Monthly Labor Review*, December 1980, pp. 3-10.

⁶ Citro, Constance and Robert T. Michael, *Measuring Poverty, A New Approach*, National Academy Press, 1995, pp. 134-140.

⁷ The questions will be in the 9th wave interview of the 1993 SIPP panel (scheduled for fielding in the period October 1995 through January 1996)

⁸ See Robert Kominski and Kathleen Short, "Developing Extended Measures of Well-Being: Minimum Income," paper presented at the Joint Statistical Meetings, Sections on Government Statistics, Social Statistics, and Survey Research Methodology, Chicago, Illinois, August 5, 1996.

“sufficient/insufficient,” and “delighted/terrible?” Are these the appropriate terms for assessing people’s attitudes and emotions about their income? What metric should be used for the scale range?

4. How easy or difficult is it for respondents to make subjective assessments of their income? How accurately do respondents review their economic situation when making subjective assessments?

To address these four main research questions, a two-prong strategy of qualitative testing consisting of cognitive interviews and focus groups is being used. The successful application of such qualitative methodologies requires that as many differing viewpoints as possible be gathered, so that a full range of ideas and opinions may be observed. In order to extend the breadth of views expressed, the study is designed to include three types of family composition, three levels of income, and five sites spread across the United States. The result is a three-by-three design matrix (based on household type and income group) with a total of nine cells. At least five interviews and one focus group are to be conducted in each cell.

Participants are screened and categorized by their household type and income level:

1. single adults with no children under 18 years-of-age in the home,
2. adults (either single or sharing expenses with another adult) with children under 18 years-of-age in the home,
3. adults sharing expenses, but with no children under 18 years-of-age in the home.

The levels of income are defined as low, medium, or high. These are determined separately for each geographic area in which testing is conducted.

The five sites are selected to provide national coverage, as well as some urban/rural differentiation. The sites are Miami, Los Angeles, Detroit, Baltimore, and West Virginia. In this paper, we only present the results from ten interviews and one focus group from one site: Miami, Florida. The Miami focus group and cognitive interviews were conducted by a research psychologist from the Bureau of Labor Statistics during the month of May 1996. Data for the other eight cells are being collected this summer and autumn.

Methodology

Our choice of two qualitative methodologies, *focus groups* and *cognitive interviews*, reflected our lack of even the most basic information concerning the ways that respondents would react to these types of questions, scales, and concepts. We had no prior hypotheses driving our data collection. Instead, we are

attempting to simply observe the procedures and strategies respondents use when making subjective assessments, hear the language they use when discussing these topics, and, ultimately, see these issues from the respondents’ perspective by encouraging and assisting them in articulating their own unique perspectives on the world.

A. Focus Groups

In our focus groups, respondents are asked to discuss the concepts of "minimum income" among themselves. Specifically, we ask them:

What amount of income would they need to "make ends meet" and how did they decide upon this amount?

What expenses would be covered by "enough money to make ends meet?" What would this amount of money purchase?

What things in their lives determine the amount of income a family needs "to make ends meet?"

In addition, participants are asked to evaluate different "levels" of income. For instance, they are asked to describe the differences between "good" and "bad" amounts of income. A similar discussion is raised for the distinctions between "sufficient" and "insufficient" incomes. Participants are asked to consider what elements or events might cause the dividing points between "good and bad" or between "sufficient and insufficient" incomes to fluctuate. Finally, participants are presented with the "delighted/terrible" scale and probed for their reaction to, as well as their understanding and use of its response categories.

B. Cognitive Interviews

The interviews begin with a concurrent think-aloud task, through which the participants are able to describe their thoughts while answering either the MIQ and an alternative "minimum spend" question. Participants are first "taught" to create an answer out loud by having the procedures explained and demonstrated. They are then requested to talk through all the issues and problems they had as they arrived at their answers to the questions. As the answers emerge, the interviewer follows up with probing questions to gain additional details and understanding. Participants are asked to rate their confidence in the answers they provided and to paraphrase some of the concepts in their own words. Such strategies are generally used to identify difficulties in understanding question wording or concepts, recall strategies, and the participants' reactions to the question.⁹

A second technique used is an "income sorting" task directed toward assessing the ways that participants evaluate income. Two versions of the task are used for each participant. In each case, cards are arranged in front of the participant that are labeled either "Very good," "Good," "Bad," "Very Bad" or "Sufficient," "Insufficient." Participants are then

handed a stack of cards with dollar amounts written on them, ranging from \$250 to \$6,000 in \$250 increments. Participants are instructed to think about all the members of their household and their expenses and then to evaluate each dollar amount as monthly take-home pay. The evaluations are made by placing each dollar amount into an appropriate category. After all the cards have been sorted and the difficult gray areas between categories recognized, the interviewers probe the decisions. Participants are asked to discuss how they decided to categorize the dollar amounts, what the income amounts categorized together had in common, and what their lives would look like with the varying categories of income.

A third task is a series of short answer questions using the "delighted/terrible" scale. "Satisfaction" assessments run the risk of being affected by preceding questions that make specific information salient, thereby creating temporary standards of comparisons, affecting judgments, and causing later responses to be higher or lower by comparison. In order to test for this possibility, a series of questions is asked for which participants are to use the "delighted/terrible" scale to identify how they felt about expenses such as the cost of feeding their families, eating out in restaurants, buying clothes, health care, transportation, school tuition, and housing. Half of the participants are asked to assess their family incomes *prior* to expenditure assessments; half of the participants are asked to evaluate their family incomes *after* the evaluation of expenses. Participants are also asked to discuss the "delighted/terrible" scale and to describe the meaning of the various categories.

In closing, the interview participants answer a series of short debriefing questions. These questions probe their reactions to the interview itself, what they liked best and least during the interview, what was easiest and most difficult, and their ideas for other questions that we could ask to more fully understand their subjective experience of the *yin* and *yang* of income and expenses.

Preliminary Findings

The preliminary findings can be categorized in terms of three main issues: ambiguity in the language, complexity in the questions, and "new" language. Each are addressed below, including examples.

A. Ambiguity in the Questions

1. Minimum Income Question (MIQ) and Minimum Spend Question (MSQ).

Half of the respondents are asked the MIQ and the other half are asked the MSQ. There do not appear to

⁹ DeMaio, T.J. & Rothgeb, J.M. (1996). "Cognitive Interviewing Techniques in the Lab and in the Field." *Answering Questions* (N. Schwarz & S. Sudman, eds., pp. 177-195). San Francisco: Jossey-Bass Publishers..

be any distinguishable differences, but the results show that both questions are plagued by the same problems.

As respondents begin to generate their answers, it becomes apparent at once that these are extremely difficult questions. In order to answer the question, it is necessary for respondents to generate a complete list of their monthly expenses and estimate a dollar amount for each expense. In many cases, respondents stop generating items and declare their list of expenses "complete," only to revise their list in response to specific probes from the interviewer or as additional items surface throughout the course of the interview. Additionally, dollar estimates generated prior to the itemized list of expenses greatly underestimate the amount of money needed to cover those expenses.

Among the focus group participants, there seems to be two distinct interpretations of the MIQ. On the one hand, about half of the participants are indicating that the question is asking for the bare minimum with which to survive-- "with no gravy," as one participant said. These participants interpret the phrase "to make ends meet" as meeting the most basic needs, such as "food on the table and shelter in which to live." For these participants, "making ends meet" is seen as equivalent to the "absolute minimum income."

A second large group of participants seems to be interpreting the question differently. For these participants, the amount of income needed to "make ends meet" includes some "gravy." This is seen most clearly when they provide a lower dollar estimate for the "*absolute*" minimum amount needed. This point of view seems to interpret "making ends meet" as starting with the current level of expenses, including some expenses that might be jettisoned if income dipped to the "absolute minimum" amount.

Most of the focus group participants seem to feel that it is relatively easy to answer this type of question because their needs are "pretty stable month-to-month." One may speculate, therefore, that if the question did not include the vague, ambiguous phrase "to make ends meet," but rather clearly defined the parameters of the judgment, then respondents might be able to provide a valid and consistent estimate. The challenge will be to convey to respondents the request for "minimum survival," while still allowing them the latitude to define what minimum survival would look like (and cost) for them.

As part of the MIQ and the MSQ, respondents are asked to consider "necessary expenses." When respondents are asked to list these expenses and explain why they consider certain items to be necessities, there is general consensus about the reasons--*they are required for existence; they are essential for survival; without them you can't live.*

Focus group participants are led through a two-step exercise of itemizing monthly expenses. First they are asked to identify those expenses that could be included as necessary for "making ends meet." They are then asked to go beyond that step and single-out only those expenses that they would deem "absolutely necessary." While there is always some diversity within groups, there is remarkable consistency across groups. It was agreed that rent, food, clothing, health insurance, transportation, utilities, household sundries, credit card payments, and garbage were considered necessary monthly expenses. Expenses for telephone, dry cleaners, and ownership of an automobile were not considered necessary. There was no consensus about expenditures for pets, school, and cable TV.

Many respondents suggested that adjustments would probably be made by changing the quantity or quality of items, rather than omitting a category completely. The lists generated from the cognitive interviews are almost identical to the focus group list. However, it is important to note that no single interview respondent generated the entire list. One may conclude, therefore, that respondents will need cues in order to make an accurate estimate of their necessary expenses. And if this is the case, this composite list may supply a good starting point for these cues.

One component of the MIQ and the MSQ is the phrase "living where you do now...." Respondents are asked to tell us in their own words what that phrase is asking them to consider. Once again, there is no consensus. The range of interpretations included *my life circumstance or stage, my place in life; my house, apartment; my neighborhood; or my city.*

One final difficulty that emerged is the interpretation of the word "you." One respondent asked the interviewer whether this question refers to the amount he personally spends or to the amount his family spends. Clearly, "you" is one of the great ambiguities in English, meaning both second person singular and plural. In order to have all respondents answering the same question, it would be absolutely necessary to clarify which of the two is being asked.

2. *The "Delighted/Terrible" Scale.*

In general, focus group participants and interview respondents seemed to like the use of emotional words in the "delighted/terrible" scale. However, when pushed to define or differentiate the terms, there is little consistency in the answers. Examples of the distinctions that are made between terms are as follows:

"Pleased" would be like "I'm making it."

"Delighted" would be like "Wow! I got a bonus!"

"Pleased" is like "good." "Mostly satisfied" is like "sufficient."

"Mostly dissatisfied" is like "OK," but "unhappy" means this job just isn't working out to give you enough money.

"Terrible" is completely hopeless.

Another problem is that the full scale of descriptors is not used. For example, some respondents viewed *mostly satisfied*, *mixed* and *mostly dissatisfied* as one group. In addition, no one reported being delighted with their income or having *terrible* income. For the most part, respondents appear to have collapsed the scale into three categories; *pleased*, *mixed* and *unhappy*.

3. Income Evaluation Questions

The question asks for judgments along two separate dimensions: "good/bad" versus "sufficient/insufficient." For some participants, these dimensions are asking for extremely different judgments. On the one hand, some participants interpreted the "sufficient/insufficient" assessment as a judgment about what is absolutely necessary for survival. On the other hand, the "very good/very bad" dimension is seen as a judgment about the "quality" of life. In some cases there is minimal overlap, between "very bad" and "insufficient," in the sense that income may be defined as "very bad" precisely because it is "insufficient." However, this comparison did not hold for the assessment that income is "sufficient" because of the conceptual dualism between bad income that is, nevertheless, "*sufficient for survival*" versus good income that is "*sufficient for some quality of life*."

B. Complexity of the Questions

1. Minimum Income Question (MIQ) and Minimum Spend Question (MSQ).

There is a clear indication that the MIQ and MSQ are complex questions to answer. The respondent must first differentiate between the current level of expenditures and the minimum needed for survival. If the "minimum expenses necessary for survival" is the preferred piece of information, the follow-up estimate of the dollar amount is somewhat a matter of factual tabulation once the respondent has decided upon which elements to include. More work needs to be done to better cue respondents and to help them answer these questions.

2. Income Evaluation Question (IEQ).

During the course of the focus group, participants are presented with a written version of the IEQ and are asked to discuss their responses. Many people reported having difficulty with this question it is difficult for some participants to conceptualize what would actually be sufficient for survival. As one respondent reported, since she had never experienced such dire straits, it is extremely "hard to decide what I need just to survive...or even less than that."

When asked to supply descriptions of what the categories "very good to very bad" and "sufficient/insufficient" might mean when applied to income, the following ideas are generated. For example, one respondent described "Very Good" income "*It is enough to be as comfortable as possible*" from "Good" income which "*It is an income that is "good" for right now, the present circumstances.*" Likewise, the conceptual distinctions between "Very Bad" income and "Bad" income: "*It means you can't pay for rent and would have to live in a shelter.*" versus "*It means there may still be hope down the road that things may be better.*" Finally, the distinctions between "sufficient" and "insufficient" income: "*It means having a little left over after you pay your monthly expenses.*" versus "*It is when you can't pay your bills and you start using your credit card to pay things off.*"

When asked which comparison they preferred, the "very good" to "very bad" judgment or the "sufficient" versus "insufficient" comparison, there is no consensus. However, there is some indication that which ever task is presented second is judged to be easier, suggesting a learning curve.

C. "New" Language

Without prompting, respondents repeatedly referred to stress, anxiety, and worry, when talking about "bad" or "insufficient" income. With good income or spending, they spoke of "freedom," "security," "not having to worry," "being more relaxed." These words came only from the participants--interviewers did not prompt them in any way. Perhaps this is a fruitful avenue for designing future subjective assessment questions. For example, we might consider these types of questions: "How often, if ever, do you worry about paying your bills on time?" "To what extent, if at all, do outstanding bills/expenses cause you concern?" "How much more money would you need each month to be free from worry about your bills?"

Conclusions and Final Comments

As the cognitive interview and focus group testing continues, we will look for consistency in our results across the five geographic areas as well as across

income groups and family types. What have we learned from this initial testing?

- It is difficult for respondents to itemize all their expenses, especially on the spur of the moment. Providing the respondent with cues and lists should improve consistency in answers.
- Respondents experience difficulty in computing the income they would "need before deductions," they know their take-home or after-deduction pay.
- When asking the IEQ , there is no consensus about the meaning of the various response categories provided by the interviewer.
- Respondents are asked to perform several tasks: (1) define the terms, (2) apply these definitions to their own lives, and (3) generate monthly estimates to convey this application. It seems that the heart of the subjective assessment is actually found in part two, the application of the categories to their unique situations. It might be feasible to supply respondents with consistent definitions so that all respondents would, in essence, be answering the same question or performing the same task.
- It may be difficult to achieve consistent answers from the D/T questions as the scale is difficult for respondents. They cope with this scale by dividing the scale into three broad groups.

In the short term, the testing described above will be completed and analyzed. In the long term, we would like to incorporate what we have learned into designing better questions and then conduct additional cognitive tests. Finally, assuming we are satisfied with the results of the cognitive testing, field testing would be required to determine reliability.

The authors would like to thank S. Fisher, D. Johnson, B. Kojetin, M. McCarthy, and C. Tucker for comments and suggestions leading to improvements in this manuscript . Thanks are also due to the University of Michigan Survey Research Center Team member, Darby Miller, to Bob Groves for field data collection and suggestions, and to the Survey Design Seminar taught by Bob Groves, Stanley Presser, and Mick Couper as part of the Joint Program in Survey Methodology at the University of Maryland. Thanks are also due to K. Short, W. Weber, and B. Passero for computations used in this paper. Finally, we wish to extend our deepest appreciation to our paper discussants, Norman Bradburn and Lee Rainwater. The responsibility for errors remains with the authors. All views expressed in this paper are those of the authors and do not reflect the views or policies of the Bureau of Labor Statistics or the views of other BLS staff members.