WHY DO RESPONDENTS PROVIDE VAGUE ANSWERS TO OPEN-ENDED, CATEGORICAL QUESTIONS?

Monica L. Dashen[1]
U.S. Bureau of Labor Statistics
Key Words: Open-Ended Questions and Vague Responses

Introduction

The goal of this paper is to determine whether respondents tailor their answers to survey questions according to their perception of an interviewer's (and/or coder's) knowledge of the survey. Often respondents who are unfamiliar with the intricacies of the survey process believe that the individual who records their responses is also the same one who examines them. If, in fact, respondents do alter their answers, ambiguity or error may result, thus affecting survey data. Respondents who perceive the interviewer (and/or coder) as knowledgeable may omit certain information to minimize redundancy or simply expedite the process. Those who perceive the interviewer (and/or coder) as less informed about the topic may be more explicit in order to compensate (e.g., Isaacs & Clark, 1987; Fussell and Krauss, 1987; Kingsbury, 1968).

If, to use an everyday example, I were asked for directions to my work place, I might say simply that it is across the street from Union Station if I deem the person asking to be familiar enough with the area to know that landmark. By contrast, if I believe the questioner to be unfamiliar with the area, I will give far more explicit directions including streets and distances.

The open-ended categorical question—one in which no category members (or response alternatives) are given—provides a good device to study the issue of response tailoring. In the absence of a list, respondents are left to their own devices as to the proper response formulation. As a result, respondents may choose to omit self-evident information to minimize redundancy. Respondents may consider question stems, such as category titles or labels, a form of self-evident information and thereby omit these titles from any responses to open-ended categorical questions. For example, when asked about any cookies purchased recently, respondents may simply say *sugar* in lieu of *sugar cookie* because the *Cookie* category title is implied. By contrast, a closed-ended categorical question provides a list, thus orienting respondents to the proper response formulation (inclusion of self-evident information in the response alternative list). Upon inspection of the response alternative list, for example, respondents might say *sugar cookie* instead of *sugar* when asked about any recent cookie purchases. For more information on the orientating effects of response alternative lists, see Schwarz (1990).

Similarly, survey respondents may try to provide the most informative response (Grice, 1975; Schwarz, 1996). In this context, "informative" refers to the omission of self-evident information, which may depend on the individual respondent's assessment of the coder's perceived level of expertise. Respondents who assume the coder is merely tabulating answers may try to be more precise or explicit in their responses (answering, for example, *sugar cookies* when asked about the category *Cookies*). By contrast, respondents who assume the coder is the same person who interviews them and designs the survey may opt to avoid redundancy (for example, answering simply *sugar*), thus jeopardizing precision in the coding process. (This assumption is reasonable given the fact that many respondents are unaware of the survey process.)

Coding precision is also often jeopardized by the interviewers' unfamiliarity with the intricacies of the coders' classification system. For example, the interviewer may fail to probe for certain information needed by the coder to correctly classify the answer (e.g., Cantor & Esposito, 1992). The omission of self-evident information may produce ambiguous answers. Such responses can cause coders, whose job it is to count and score all responses, to disagree about the respondent's intent or the "true" meaning of the response and thus lead to a degradation in the coding process. That is to say, coders may follow their own informal rules, which are not uniformly applied by all coders, to accommodate those "hard to code" responses. As a result, these informal rules may reduce coder agreement (Conrad and Couper, 2001).

It may seem reasonable to suppose that if respondents routinely try to make their responses as informative as possible by omitting category titles (self-evident information) from their responses to open-ended questions, ambiguous responses should therefore be easy to interpret. All one need do is add the category title to the response. This solution, however, is not always viable. Evidence by Dashen & Fricker (2001) indicates that people tend to formulate goal-oriented criteria of inclusion for open-ended categorical questions, two of the most common types of which are "to make" and "to accompany." When asked about *Cookie* purchases, people using the "to make" criterion may say *sugar*, *water*, and *chocolate*. Respondents using the "to accompany" criterion may respond with *ice cream*, *milk*, or *plate*.[2]

In summary, respondents may omit self-evident information (category title) to open-ended categorical questions

[1] The views expressed here are those of the author alone and do not necessarily reflect the official position of the Bureau of Labor Statistics. The author assumes responsibility for any inaccuracies and would like to thank Fred Conrad, John Dixon, Jim Esposito, and Bill Mockovak for their insightful comments. I would also like to thank Pam Douglas and Sarah Lewandowski for their assistance in the collection and coding of the data.

[2] The tendency to include items that do not belong to the category is further compounded by the fact that the categorical distinctions of these questions are not always what a reasonable respondent would expect. For example, the Telephone Point of Purchase Survey (TPOPS), which tracks consumer spending habits, does not include *cookies* and other baked goods such as *cakes* and *pies* in the same category.

depending upon their assessment of the coder's level of knowledge. If the respondents believes that the coder is knowledgeable, the respondents may be more likely to omit self-evident information from their responses than if they believe that the coder is not veryknowledgable. The omission of the self-evident information may produce ambiguous responses that are difficult to code and may subsequently affect data quality.

The present work has two specific aims. The first is to determine whether people tailor their responses to open-ended categorical questions based on their understanding of the coder's level of knowledge. Respondents were given category titles under one of two conditions: Best- and Worst-Case scenarios. In the Best-Case scenario, the respondents were told that the coder was an expert familiar with the survey goals. The respondents in the Worst-Case scenario were told that the coder was a novice unfamiliar with the survey goals. Respondents in both conditions were asked to generate items they thought belonged to the categories. Afterwards, for each item generated, all respondents were asked to provide additional information about the inclusion or exclusion of the category title. This task (category title inclusion justification) serves to provide insight into the possibility that people deliberately omitted the category title from their response based on their understanding of the coder's knowledge.

The second aim is to replicate those findings of Dashen and Fricker (2001) that demonstrate that people systematically formulate goal-oriented and literal criteria of inclusion for open-ended categorical questions. Under the goal-oriented interpretation, when asked about *Bread*, people may include *butter*, *knives* and *jam* as things that "accompany" *Bread.* Or they may say *flour*, *salt* and *baking soda* and justify them as all things that are needed "to make" bread. Under the literal-interpretation, when asked about *Bread* people may include *butter topped bread*, *white bread* and *rye bread* and justify them as types of bread. The exemplar generation task and item inclusion justification serve to detect the respondents' criteria of inclusion in the present work. A similar methodology was used in this previous work.

The present work has practical applications. A conscientious survey designer strives to identify and prevent coder disagreement before it occurs. It is reasonable to assume that a closer fit between the survey designer's desired response format and the respondent's response format will reduce coder disagreement because there should be fewer ambiguous responses. For that reason, the present work suggests that the survey designer should make it a point to specify the desired response format prior to the onset of the survey

Study

Methods

Participants

Thirty-eight participants (16 males and 22 females) responded to an advertisement in a local newspaper and received $35.00 each to compensate for their participation. The participants' mean age was 37.9, and their average educational level was 15 years of schooling (or 3 years of college).

Procedure and Materials

Respondents participated in three phases. The phases are described below.

Phase 1: Scenarios. All respondents were told about the coder's job duties (e.g., counting and scoring all responses in a timely manner). Of the 38 respondents, 19 were randomly assigned to the Best-Case scenario and 19 to the Worst-Case scenario. Those respondents in the Best-Case scenario were told that the coder had been on the job for 10 years and was familiar with the survey goals and types of responses provided. In contrast, those respondents in the Worst-Case scenario were told that the coder had been on the job for less than one year and was unfamiliar with the survey goals and types of responses provided.

Phase 2: Exemplar Generation Task. Respondents in both groups were asked to engage in an exemplar generation task, in which they were given a series of open-ended categorical questions. Each participant received a booklet containing instructions and nine category titles. Each category question was on a separate page with ample space for participants to write down all relevant purchases. The participants were required to generate example purchases for the following categories: (1) *Bread*, (2) *Breakfast Cereal*, (3) *Tomatoes*, (4) *Cookies*, (5) *Lettuce*, (6) *Potatoes*, (7) *Apples, (8) Soups, and (9) Bananas.* Though all participants saw the same set of category titles, no two people saw the same order of category titles in the booklet. (Note: These categories were patterned after the Telephone Point of Purchase Survey questions (TPOPS).) All categories were designed in such a way that only literal instantiations belonged in the category. For example, the *Cookies* category consists of items such as *sugar cookie* and *chocolate chip cookie*. Similarly, the *Bread* category consists of such items as *whole wheat bread, French bread*, and *rye bread*.

The respondents were instructed to interpret the open-ended categorical questions as hypothetical. For example, one of the questions read: "Hypothetically, if you were to have made a purchase from the category *Bread* within the last two weeks, what items would you have purchased? Write down each purchase in the space provided below." Thus, respondents were not limited by their actual purchases in listing items for the categories. In addition, respondents were encouraged to say more than one item for each category.

In order to emulate a telephone survey, respondents were instructed to complete the survey in a sequential order. The respondents were not allowed to skip ahead to questions, nor were they allowed to return to previously answered questions. The experimenter monitored the respondents while completing the task to insure that they complied with the instructions. In keeping with the TPOPS methodology,

respondents were not told that the same answer could not be used for two different categories.

Phase 3: Justification Task. After completing the exemplar generation task, respondents were asked to justify their responses in two ways. First, respondents were asked to say why they included the item in the category (item inclusion justification). This form of justification provides insight into how the respondents interpreted the category title. Second, respondents were asked to state why the category title was included or not included in their response (title inclusion justification). This form of justification provides insight into whether respondents were deliberately omitting the category titles from their responses. It should be noted that people were given an unlimited amount of time to generate exemplars and justifications.

Results

The discussion of the data analysis has been broken down into two sections. The first section describes the scoring procedure for the exemplar generation and justification tasks, and the second section discusses the results of these procedures.

Description of Scoring Procedure

The scoring procedure for the exemplar generation task had two parts. First, the fictitious purchases listed were scored as intended, unintended, or ambiguous reports, based on whether they corresponded to the intentions of the designers of the TPOPS survey. Second, the two types of justifications were collected and classified into various categories for further analyses. These two procedures are further discussed in the following two sub-sections.

Scoring of Listed Fictitious Purchases. There was a total of 1839 items recorded for all nine categories across all respondents. The average number of items recorded per category across respondents was 204.33 (1839/9). Thus, each person reported an average of 5.4 items per category. Because some categories are more broadly defined than others, some people might have written down more items for one category than other categories. The number of items per category ranged from 109 items--assigned to the *Bananas* category, to 362 items--assigned to the *Bread* category.

For each participant, the items or fictitious purchases reported for each category were classified into three categories: (a) intended; (b) unintended; and (c) ambiguous exemplars. Definitions and examples of these categories are provided in Table 1 below.

Table 1 Exemplar Type by Definition and Example

Exemplar Type *(col. 1)*	Definition *(col. 2)*	Examples *(col. 3)*
Intended Exemplars	Intended exemplars are those items that are seen on the TPOPS list in the desired format. These items do not have an ambiguous meaning.	(1) *Sugar cookies* for *Cookies*, (2) *Cherry tomatoes* for *Tomatoes* and, (3) *Rye bread* for *Bread.*
Unintended Exemplars	Unintended exemplars are those items that are not seen on the TPOPS list. These items do not have an ambiguous meaning.	(1) *Spoon* for *Soup*, (2) *Ice cream* for *Banana*, and (3) *Pan* for *Cookies*.
Ambiguous Exemplars	Ambiguous exemplars are those items that are not seen on the TPOPS list. These items do have an ambiguous meaning.	(1) *Cherry* for *Tomatoes,* (2) *Sugar* for *Cookies, and* (3) *Sandwich* for *Bread***.**

As can be seen in column 2 of Table 1, intended exemplars are those items that are on the TPOPS list and only have one meaning. These items are literal instantiations of a particular category, for example, *sugar cookies* for *Cookies*. The definitions of the unintended exemplars differ from the definitions of the intended exemplars in that the items are not seen on the list.[3] The ambiguous exemplar definitions differ from the intended and unintended exemplars in that the ambiguous exemplars have more than one meaning. (For more details, see Dashen and Fricker 2001)

Scoring of Justifications. After completing the exemplar generation task, respondents were asked to provide two types of justifications for each item listed. For the first justification, item inclusion, responses to the question, "Why do you think this item is a member of the category?" in the exemplar generation task were collected and scored. This justification was designed to detect the respondent's criterion of inclusion. Responses to this item inclusion justification were classified into one of four major groups: (a) literal (type); (b) to make; (c) to accompany; and (d) does not make sense/no justification provided. First, the "literal" group included participants who interpreted (or justified) the category titles in a literal and narrow manner. That is to say, respondents tended to comment on the fact that it is an instantiation of a category (e.g., "golden delicious apples are a type of apple"). Second, the "to make" group involved those respondents who justified their responses as an ingredient (e.g., "apples are used to make apple cider"). Third, the " to accompany" group included those participants who said that the item was used to accompany the category (e.g., "cinnamon is a topping for apples").

It becomes necessary at this point to review the procedure for developing the coding scheme of the item-inclusion justifications. After inspecting the justifications, I

developed the above-mentioned coding scheme. Using this coding scheme, two judges who were blind to the nature of the study classified all responses into four major groups. The judges' response classifications were correlated at .94, as a means of estimating reliability. There was a total of 1839 reports. As an additional measure of agreement between the two judges, a Kappa was computed to correct for chance between raters. The Kappa yielded a value of .92. In all cases, the first judge's codes were maintained for the justification analyses reported in the next section.

For the second justification, title inclusion, responses to the question, "Why did you include the category name (or not) in each one of your answers?" in the exemplar generation task were collected and scored. These justifications served to detect the respondent's criterion for response formulation and were classified into one of three groups: (a) to be more specific; (b) it was implied; and (c) did not make sense/ no explanation was provided. The first category, "to be more specific," included those participants who said that they included the category title when listing its contents as a means of being more specific. For example, a respondent included "bread" when listing "sandwich bread" as a type of bread. The second category, "it was implied," involved participants who did not say the category title when listing its contents because they found it unnecessary or redundant. For example, when listing the contents of the *Bread* category, these people only said "sandwich" rather than "sandwich bread" and justified their failure to say the "bread" as something the coder already knew. The remaining category, "does not make sense/no justification" included those respondents who simply failed to provide an explanation.

The procedure for developing the coding scheme for the title inclusion justifications was as follows. After inspecting the justifications, the author developed the above-mentioned coding scheme. Using this coding scheme, two judges who were blind to the nature of the study classified all responses into 3 major groups. The response classifications of the two judges were correlated at .97, as a means of estimating reliability. There was a total of 1839 reports. As an additional measure of agreement between the two judges a Kappa was computed to correct for chance between raters. The Kappa yielded an identical value (K = .95). In all cases, the first judge's codes were maintained for the justification analyses reported below.

Exemplar Generation Task Performance.

There are four questions to be answered. (a) Are there some interpretations that lead people to produce correct answers (intended exemplars)?; (b) Are there some interpretations that lead people astray?; (c) Is there a response formulation that leads to ambiguous responses?; and (d) If so, is this formulation contingent upon a respondent's perception of the coder's level of knowledge? The first two questions serve to replicate earlier findings by Dashen and Fricker (2001), by focusing on why people produce correct or incorrect (false report) answers through an examination of the criterion of inclusion. In contrast, the latter two questions focused on response tailoring and the conditions under which it occurs.

The percentages of responses classified within each of the four general groups of justifications ("literal," "to accompany" goal, "to make" goal, and "no justification") were calculated by exemplar type; these results, which are summarized in Table 2, help to answer the first two questions.

Table 2. Exemplar Type by Item Inclusion Justifications.

	Item Inclusion Justifications		
	Literal	Accompany	To Make
Exemplar Type	*(col..2)*	*(col..3)*	*(col..4)*
Intended	97% (278/287)	0% (0/287)	0% (0/287)
Unintended	17% (111/668)	42% (282/668)	39% (263/668)
Ambiguous	62% (546/884)	10%(89/884)	24%(210/884)

The goal of the first question was to find out which method people use when producing the correct answer (or intended exemplar). The frequencies of intended exemplars (as contained in the intended exemplar entries for each justification listed in Table 2, columns 2-5) suggest that when people adopted the literal list method, they were most likely to generate intended exemplars, $x^2(1) = 252.13$, $p = .00$. This finding replicates those of Dashen and Fricker (2001).

The goal of the second question was to find out what method led people astray when they produced unintended exemplars. The frequencies of the unintended exemplars (as contained in the entries for each justification of Table 2, columns 2-5) suggest that when people adopted either the "accompany" or "to make" goal-oriented method, they were most likely to generate unintended exemplars $x^2(3) = 297.02$, $p = .00$. This finding also replicates those of Dashen and Fricker (2001). (The percentage of "no justifications" was also included in the above two chi-square analyses.)

In summary, the literal method of interpretation led to correct answers, whereas the "accompany" and "to make" methods did not. Interestingly, many of the interpretations, particularly those that employed the literal method, led to ambiguous responses. The remaining two questions focused on the origins of the ambiguous responses (or omission of the category title).

The goal of the third question was to discover, which interpretation produced the most ambiguous responses. The frequencies of ambiguous exemplars (as contained in the ambiguous entries for each justification of Table 2, columns 2-5) suggest that when people adopted the literal method they generated the more ambiguous exemplars than when they adopted any of the other interpretations, $x^2(1) = 707.21$, $p = .00$.

To further understand why ambiguous responses were produced with the literal interpretation, I examined the title inclusion justifications for intended and ambiguous exemplars that occurred when people used this method.

Table 3.Title Inclusion Justifications by Exemplar Type

	Exemplar Type	
	Intended	Ambiguous
Title Inclusion Justifications	*(col. 2)*	*(col. 3)*
Specific	96% (275/287)	.007% (6/884)
Implied	.01% (3/287)	95% (839/884)
No. Just.	.03% (9/287)	.04% (39/884)

As can be seen in column 2 of Table 3, respondents explained their inclusion of the category title in intended exemplars with the justification "to be specific" (96% or 275/287) more often than the justification of "implied" (.01% or 3/287), $x^2(2) = 504.45$, $p = .00$. In contrast, as can be seen in column 3 of Table 3, respondents tended to explain their exclusion of the category title in ambiguous exemplars with the justification of "implied" (95% or 839/884) more than with the justification of "to be specific" (.007% or 6/884), $x^2(2) = 1510.16$, $p = .00$. (The percentage of "no justifications" was also included in the above two chi-square analyses.)

The goal of the fourth question was to build upon the previous results by focusing on the possibility that people tailored their answers (i.e., omitted the category title) based on their understanding of the coder's level of knowledge. As expected, Best-Case Scenario respondents were more likely to omit the title from their literal instantiation answers (53%) than Worst-Case Scenario respondents, (69%), $x^2(1) = 22.16$, $p = .00$.

A further analysis was conducted to find out whether exemplar frequency was a factor in response tailoring (i.e., omission of the category title). Exemplar frequency refers to how many exemplars were generated for each category in the exemplar generation task. Respondents may have been more likely to omit the category title for frequent exemplars than for infrequent ones because they may have believed that the coder was already familiar with the exemplars. In contrast, respondents may have included the category title for infrequent exemplars due to their unfamiliarity. To address this exemplar frequency issue, I calculated the frequency of each exemplar per category in order to estimate the frequency of each exemplar (e.g., *raisons* were seen five times in the *Bread* category, while *short* was only seen once in the *Bread* category).

Table 4 summarizes the proportions of high and low frequency ambiguous exemplars that used the item inclusion justification (literal) for each scenario.

Table 4. Ambiguous Exemplar Frequency (under the literal justification) by Scenario

	Ambiguous Exemplar Frequency	
Scenario	High *(col. 2)*	Low *(col.3)*
Best-Case	61% (208/344)	59%(120/202)
Worst-Case	40% (136/344)	41% (82/202)
Totals	*63% (344/546)*	*37% (202/546)*

There are two points to be made about the above table. First, exemplar frequency appeared to play a role in response tailoring. Respondents in both scenarios were more likely to omit the category title for high frequency exemplars (63%) than for the low frequency exemplars (37%), $x^2(1) = 36.93$, $p < .001$. Second, it is quite possible that people in the Best-Case Scenario group simply omitted the title due to exemplar frequency rather than coder expertise. If this were the case, we would expect to see no differences between the two scenarios for each level of exemplar frequency. However, this was not the case. Respondents in the Best-Case Scenario group were more likely to omit the category title for high frequency exemplars (61% or 208/344) than those in the Worst-Case Scenario group (40% or 136/344), $x^2(1) = 15.07$, $p < .001$. Likewise, respondents in the Best-Case Scenario group were more likely to omit the category title for low frequency exemplars (59% or 120/202) than those experiencing Worst-Case Scenario (41% or 82/202), $x^2(1) = 7.15$, $p < .001$.

Conclusions

This study's exemplar generation and justification data addressed a theoretical issue that is central to survey methodology: Do respondents tailor their answers to the coder's level of knowledge? In this study, the respondents were given a category title (e.g., *Cookies*). They were then asked to list items they thought belonged in the category and to justify those items under one of two conditions: Best-Case Scenario or Worst-Case Scenario. While Best-Case Scenario respondents were told that the coder was very experienced with and knowledgeable of the survey goals, the Worst-Case Scenario respondents were told otherwise. For example, a respondent in the Worst-Case Scenario could say that *sugar* belongs in the *Cookie* category and justify that response by saying that it is type of cookie and the title is implied. (Such a justification would be classified as "a literal instantiation" and "implied" since the respondent's justification was that the item is a type of cookie.)

Best-Case Scenario respondents were more likely to tailor responses (i.e., omit category titles) than their counterparts in the Worst-Case Scenario. Furthermore, all respondents were more likely to omit the category title from frequent exemplars

compared to infrequent exemplars generated in the exemplar generation task.

The results also served to replicate the criterion inclusion findings of Dashen and Fricker(2001). Respondents were most likely to provide intended exemplars when they adopted the literal strategy and provided unintended exemplars when they adopted the "accompany" or "to make" justification.

The present work has important practical implications. Conscientious survey designers are interested in obtaining codable responses. The results of the present work strongly suggest that category titles alone do not necessarily attain that end. What may be an obvious formulation to designers may not be obvious to respondents. In other words, unless category titles are supplemented with some sort of instructions about how responses should be formulated, survey designers are almost certain not to get consistent results. Survey designers can provide a lead-in statement to help respondents understand a category's desired method of formulation. Such a statement needs to take into account the nature of the category. For example, a category defined simply by instantiations of the title (e.g., *Bread*) would be better served by an instruction telling respondents to include the category title. On the other hand, a category that the TPOPS survey designers have constructed to include more than simple instantiations of the title (e.g., the computers category, which includes components) requires broader instructions about how to formulate a response.

The present work is a step toward understanding how people formulate responses to open-ended categorical questions. One fruitful avenue of future research may involve examining the role of satisficing in response tailoring (Krosnic,1991). Another fruitful avenue for future research involves examining how category similarity influences respondents' interpretations.

References

Cantor, D. & Esposito, J. (1992). Evaluating Interviewer Style for Collecting Industry and Occupation Information. American Statistical Association *Proceedings of the Section on Survey Research Methods*, 661-666.

Conrad, F.G., and Couper, M.P., (2001) Classifying Open-Ended Reports: Coding Occupation in the Current Population Survey. Paper presented at the *Federal Committee on Statistical Methodology (FCSM) Research Conference*, Arlington, VA, November.

Dashen, M. & Fricker, S. (2001). Understanding the Cognitive Processes of Open-Ended Categorical Questions and their Effects on Data Quality. *Journal of Official Statistics*. 17(4) pp. 457-77.

Fussell, S.M. & Krauss, R. M. (1989). The Effects of Intended Audience on Message Production and Comprehension: Reference in a Common Ground Framework. *Journal of Experimental Social Psychology*. 25, 203-219.

Grice, H.P., (1975). Logic and Conversation. In P. Cole & J.L. Morgan (Eds.). *Syntax and Semantics* (Vol.3). New York: Academic Press.

Isaacs,E.A., & Clark, H., H., (1987). References in Conversation between Experts and Novices. *Journal of Experimental Psychology: General*, 116, 26-27.

Kingsbury D. (1968). Unpublished Honors Thesis, Department of Social Relations, Harvard University.

Krosnick, J. A. (1991). Response strategies for coping with cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*. 5, 213-236.

Schwarz, N. (1990). Assessing Frequency Reports of Mundane Behaviors: Contributions of Cognitive Psychology to Question Construction. (pp.98-119) In C. Hendrick & M.S. Clark (Eds.), *Research Methods in Personality and Social Psychology*.

Schwarz, N. (1996). Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Telephone Point of Purchase Survey (1997). Bureau of Labor Statistics Handbook of Methods. Bureau of Labor Statistics.