

CATEGORIZATION OF RESPONSES TO THE OPEN-ENDED LABOR FORCE QUESTIONS IN THE
CURRENT POPULATION SURVEY (CPS)

Maria P. Fracasso, U.S. Bureau of Labor Statistics
441 G St, NW, Washington, DC 20212

KEY WORDS: Misclassification, Response Error

The Current Population Survey (CPS) is a monthly household survey that determines, among various other estimates, the current unemployment rate. The survey makes use of a structured questionnaire which elicits information on individuals' labor force activity. Although continuous modification and expansion of this structured questionnaire has occurred since its inception in 1940, recently there has been increasing interest in studying specific areas of the questionnaire such as the categorization of open-ended questions. The reason for the focus on categorization is due to the fact that many questions in the CPS which establish respondents' labor market activity, including labor force status, during the survey reference period are open-ended.

When researchers obtain a set of qualitative materials, their first goal is to classify the content into appropriate categories in order to describe it systematically. Hence, the purpose of categories for open-ended questions is to organize a great many items into a small number of classes for the situation under investigation to be more easily understood. Open questioning is a technique designed to be as close as possible to a natural conversation. The open-ended question does not designate specifically to the respondent the possible answers to the question. Instead, the interviewer is expected to record the respondents' answers verbatim. One of the ways to convert responses to an analyzable form is conducted through field coding. Field coding consists of having the interviewer interpret respondents' answers and fit them into a system of fixed response categories. However, there has been considerable concern over the difficulty in processing responses to field coded, open-ended questions in a reliable way.

Consistent coding of qualitative material may be influenced by response categories which are inappropriate, nonexhaustive or not mutually exclusive. Therefore, in any application of open-ended questions, an explicit and objective scheme of categorization which makes possible the consistent and precise coding of qualitative materials must be developed in order to achieve the goal of accurate quantification and measurement. As suggested by Lazarsfield and Barton (1955), the primary way to start formulating response categories is with the sort of concrete categories which experienced policy-makers or participants in the situation use, worked out in as clear and logical form as possible. Furthermore, to figure out what theoretical categories will be applicable to a given field of behavior, it is necessary to continuously refine concrete categories closely adapted to the data themselves. Finally, the merit of adding a new category is assessed by determining whether the new category would be meaningful within the rationale of the system of categories involved by ascertaining whether the

example could not just as well be placed in an already existing category and by judging whether the new category would be used frequently enough to warrant its separate designation outside an "other" category.

With respect to the CPS, five of the seven field coded, open-ended questions involved in labor market analysis were examined in this research. Recently, there has been concern over whether some of the categories for these questions were obsolete and, hence, were used infrequently or not at all. Furthermore, there was increasing question that if categories were being used infrequently, then a substantial proportion of responses were coded as "other." Placing a sizable number of responses into the "other" category not only creates the problem of inaccuracies in the classification of an individual's employment status, but also decreases the detail of demographic and economic characteristics. The frequent use of an "other" category for classification of open-ended questions suggests that additional or alternative categories may be essential for reliable, consistent data.

Consequently, the focus of this study was threefold: 1) to determine if current response categories for the open-ended CPS questions are appropriate, necessary, sufficient, and hence useful for labor force classification; if not, 2) to develop an alternative set of response categories for field coded, open-ended questions in the CPS, and finally, 3) to determine whether categorization of responses are both accurate and consistent in coding responses into individual categories and major labor force classification groups (employed, unemployed, not in the labor force) as a consequence of the set of response categories.

METHOD

Initially, the frequency of use for each of the present response categories for the open-ended CPS labor force questions was reviewed. Then, the types of responses that were consistently and frequently coded as "other" was observed. An alternative set of categories was developed by eliminating those categories that were used very infrequently within the last year and by adding new categories based on frequency of specific responses that were consistently placed in the "other" category as well as an unstructured classification of these responses by CPS analysts. Finally, the reliability of category versions, for coding responses into each category for accurate classification, was determined. Reliability of categories was measured using a response sorting technique. A set of sample responses was derived for each of the open-ended questions from a review of numerous actually transcribed responses. These responses were then sorted into categories which were thought to be most appropriate, using either the present or alternative versions.

Development of sample responses

For this study, sample responses for each of the open-ended questions were reviewed from three separate sources. Transcribed sample responses were obtained: 1) for all of the categories to the CPS question 20C "What is the reason...worked less than 35 hours LAST WEEK?" and "What is the reason...USUALLY works less than 35 hours a week?" 2) for the "other" category in all the CPS open-ended questions, and for all the categories to the labor force questions in the National Longitudinal Survey (NLS). The NLS is an ongoing survey of individuals throughout the country which obtains personal information including labor market activity. Labor force status and activity are obtained by asking the identical CPS questions, including the open-ended labor force questions.

The transcribed responses from these sources served as the pool of sample responses for their respective questions. A total of fifty responses for each question were selected for use in the study and both responses that were "easy" (i.e. had been coded into an existing category) and "difficult" (i.e. had been coded into the "other" category) were selected. The number was determined as an attempt to decrease the burden placed on the FR's who had volunteered to participate, yet to represent the type of responses that were used for all different categories for each of the respective questions.

Development of an Alternative Set of Categories

The development of an alternative version of response categories was accomplished using two separate methods. The first method used an adaptation of a procedure using by Schumann & Presser (1981). This procedure examined previous category use for all of the CPS open-ended questions during the 1988 calendar year as well as type and frequency of responses placed into the "other" category during three months in 1988.

Results of the tabulation of frequencies for each of the categories to the open CPS labor force questions for 1988 indicated, that for some questions, several categories were used infrequently. In addition to the apparent underuse of many categories, there seemed to be a substantial proportion of responses placed into an "other" category. The frequent use of this catch-all category to classify responses indicates that choices may be too limited to accommodate the responses obtained.

The second method consisted of soliciting a group of CPS analysts to sort, for all of the five different questions, each of the sample responses into what they believed to be categories consisting of similar responses. A total of seven CPS analysts, employees of the BLS, participated in this portion of the study. These participants were instructed to develop their own categories. Findings of the classification of responses into "like" groups supported the development of an alternative set of categories per question.

As a consequence of both methods, an alternative set of response categories was developed for each of the CPS open-ended questions. For some questions, categories were

added to accommodate responses that appeared frequently in the "other" category. Additionally, for some questions, category labels were changed in order to more accurately describe the kinds of responses that belonged in the appropriate category. Finally, some of the present categories were collapsed when it was indicated that they were used very infrequently.

Reliability of Field Representative's (FR) Response Categorization

A total of 128 CPS FR's from six regional offices participated in this portion of the study. Categorization of the 50 sample responses for each of the five CPS open-ended questions was completed by each CPS FR. All participants were asked to read each of the 50 sample responses and to place it into one of the categories that were provided for the respective question. Half of the group was asked to classify the responses using the present CPS categories and the remaining half was asked to classify the SAME responses using the alternative categories. All participants worked individually at sorting the responses. FR's were NOT given any instructions on the use of the alternative categories, nor were they allowed to refer to the FR's manual in using the present categories.

RESULTS

The results may be seen as somewhat restrictive in that the responses chosen for this study are simply examples and do not necessarily reflect either the identical frequency with which individuals provide these responses or the entire universe of actual responses to the CPS labor force questions.

Development of Alternative Response Categories

20C(1) What is the reason...worked less than 35 hours LAST WEEK?

In an attempt to clarify category membership of certain responses there was a modification of the present categories "bad weather," "own illness," and "on vacation" to "bad weather affected job," "illness, injury, medical appointment," and "vacation/personal day," respectively. The categories "plant and machine repair," and "material shortage" were eliminated because of their infrequent utilization and the category "slack work" was expanded to include these responses and relabelled "slack work/business conditions." In addition, the categories "job started during week" and "job terminated during week" were collapsed because of infrequent use. The category "too busy with housework, school, personal business" was expanded into two categories, "family/personal obligations" and "school responsibilities" to provide additional information for labor market analysis. Finally, the category "hours vary each week" was developed to accommodate responses that indicated changes to the hours scheduled and flexible hours, and "seasonal" was constructed to incorporate responses which indicated the respondent was working at a seasonal job.

20C(2) What is the reason...USUALLY works less than 35 hours a week?

In order to obtain more precise coding of responses, the present category label "own illness" was modified to "health/ medical limitations." As in the previous question, the category "too busy..." was replaced with "family/personal obligations" and "school responsibilities." An additional category was added, "childcare problems," to retrieve more detailed information. After reviewing the frequency with which the category "full-time work week less than 35 hours" was used and the type of responses which were placed into this category, it seemed apparent that two categories were needed to reduce coding error. One category was developed to accommodate responses indicating that the respondent's work week hours were limited by regulation, "hours restricted by regulation," and the second category included responses which suggested that the respondent just worked less than 35 hours in a work week, "FT work week less than 35 hours."

21A Why was...ABSENT from work last week?

The categories of "own illness," "on vacation," and "bad weather" were changed to "illness, injury, medical problems," "vacation/personal day," and "bad weather affected job." Since the categories "temporary layoff" and "indefinite layoff" were seemingly used interchangeably, they were collapsed into only one category of "on layoff." A number of new categories were developed, "slack work," "maternity/ paternity leave," "family/ personal obligations," "civil/ military duty," and "school responsibilities," as a consequence of the frequent placement of such responses into the "other" category.

22A What has...been doing in the last 4 weeks to FIND WORK?

To sort out the responses that were considered passive job search methods into individual and separate categories, "looked at ads" and "attended job training programs/ courses" were developed. Furthermore, the categories, "registered or contacted: union/ professional organizations; school/ university employment centers," were added to accommodate frequent "other" responses.

22E Could...have TAKEN A JOB last week if one had been offered? NO: Why not?

Presently, this question consisted of very restricted response categories with numerous responses clustered into the "other" category and consequently confining the categorization of many responses which were given to this question. Review of the "other" category supported the construction of two new categories, "family/home responsibilities" and "child care problems."

Effect of Alternative Categories on Coding Error

FR's categorization of the 50 sample responses to each of the five open-ended labor force questions was collected and the proportion of accurately classified responses per individual category was computed for each question in both versions. For each of the categories per

question, t-tests were computed comparing alternative and present categories for the frequency of accurate response classification. The results reflect the differences between the alternative and present version of response categories to accurately code responses.

20C(1) What is the reason...worked less than 35 hours LAST WEEK?

With one exception, the alternative categories either resulted in significantly higher correct coding rates or in no difference in correct coding rates. Consequently, changes allowed coding of the responses more consistently and accurately. The exception consisted between the comparison of the alternative, "job started/ ended during week" and the present category "job terminated during week" where the present category had a significantly higher correct coding rate. Review indicated there was a problem in distinguishing whether the responses were a result of job termination or "business conditions" (which is the label of a separate category on the alternative form). Therefore, without explicit definitions there may be difficulty in categorizing.

20C(2) What is the reason...USUALLY works less than 35 hours a week?

As with the previous question on hours worked, the alternative categories either resulted in significantly higher correct coding rates or in no differences with the exception of one category. The exception occurred when comparisons were made between the new category "childcare" of the alternative set and the "too busy..." of the present set. Review of the categorization revealed that the alternatives allowed two categories "family/ personal obligations" and "childcare" for which similar responses could be placed. The separation of this category, without explicit training on category usage, may cause confusion.

21A Why was...ABSENT from work last week?

All of the alternative categories yielded significantly higher or no difference in correct coding rates. Modified labels, collapsed categories and additional categories provided sufficient amounts of information for significantly more accurate and consistent coding of appropriate responses.

22A What has...been doing in the last 4 week to FIND WORK?

All of the new alternative categories resulted in significantly higher coding rates. Review of the categorization with the present set revealed that responses for each of these categories were seldom placed into the correct category and, more importantly, were being arbitrarily placed across all different response categories. More explicit, individual and separate response categories promoted more frequent correct coding of responses.

22E Could...have TAKEN A JOB last week if one had been offered? NO: Why not?

For both new categories of the alternative set there were significantly lower correct coding rates. Review of the categorization revealed

that the more consistent coding with the present set was a consequence of the limited number of categories available for coding and hence, the relatively large number of responses that were expected to be put into the "other" category. In addition, it appeared that the appropriate categorization of responses was not clear. This was especially problematic when the categories were as similar and as easily confused as "family/ home responsibilities" and "child care problems." The objective for adding these categories was to reduce the frequency of "other" category usage and to increase detailed information about the respondents. Results indicated that, with these responses, the "other" category WAS reduced with the alternative form.

Effect of Alternative Categories on Labor Force Classification

All of the field coded open-ended labor force questions use the response categories to classify the respondent into two or more groups (e.g. employed/ unemployed). For each of the questions, comparisons were made of the proportion of responses expected/ observed to fall into the specific major classification grouping between the present and alternative version of response categories.

20C(1) What is the reason...worked less than 35 hours LAST WEEK?

The purpose of this question is to distinguish individuals who worked part-time hours (<35 hrs) for ECONOMIC reasons, such as "slack work/ business conditions" or for NONECONOMIC reasons, such as taking care of family obligations or vacation. There were no significant differences between those responses which were expected and observed as part-time for economic and noneconomic reasons between the alternative and present response categories.

20C(2) What is the reason...USUALLY works less than 35 hours a week?

This question differentiates part-time work for economic or noneconomic reasons. Significant differences were found between versions, with the present set more accurately categorizing responses into the part-time for economic reasons classification. The reason for the lower consistency for the alternative category coding appeared to be related to the collapsed category "slack work/ business conditions." Without clarification concerning the appropriate usage for this category, there was confusion and less consistent coding of these responses. However, for the coding of responses into categories designating part-time for noneconomic reasons, there was more consistent sorting of these responses with the alternative than with the present categories. Results suggest that with the new categories, there is more consistent coding of responses into the part-time for noneconomic reasons, thus reducing classification error.

21A Why was...ABSENT from work last week?

This question discriminates between individuals who were absent from work (inactive) and have a job (employed) from those who were absent and don't have a job (unemployed).

Whereas there were no significant differences in coding the unemployed reasons, significant differences were found between versions for the coding of responses into the inactive-employed group. The alternative categories promoted significantly higher classification rates for the inactive employed responses than the present categories.

22A What has...been doing in the last 4 week to FIND WORK?

This question distinguishes active and passive methods of job search with the goal of determining which individuals were actively looking for work and are, therefore, unemployed (assuming availability for work) and those who were passively looking and are, therefore, not in the labor force. No significant differences were found between versions for the responses coded as active methods of search. However, significant differences were revealed for the responses which were to be coded as passive methods of job search. There was more accurate and consistent coding of responses for the not in the labor force classification using the alternative rather than the present response categories. Data indicated that with the responses coded as "other" under the current categories were coded into passive categories under the alternatives. This change in categorization from "other" to a new category specifying a passive category altered the respondents classification, and with the alternatives, there was more accurate categorization. This modification, however, does not always imply a change in classification since classification is based on all job search activities conducted during the specified four week period.

22E Could...have TAKEN A JOB last week if one had been offered? NO: Why not?

This question determines the individual's availability to work. Responses to this question are sorted into categories which result in classifying the respondent as either unemployed (for reasons such as temporary illness) or not in the labor force (for reasons such as going to school). While no significant differences were revealed between versions for classifying the respondent as unemployed, significant differences were found in coding responses which would place the respondent into the not in the labor force classification. There was more accurate coding of responses which classified the individual as not in the labor force with the alternative than the present categories. The addition of these categories provide supplemental information about the respondents who indicate they are unavailable for work as well as decrease classification error.

DISCUSSION

Following a review of the current category usage for the field coded, open-ended CPS labor force questions, an alternative set of response categories was developed. Categories were adapted from actual data in that they were constructed based on respondents' answers to these CPS questions as well as CPS analysts' recommendations for new categories. Categories were added or modified when it was understood that these alternatives were meaningful within the rationale of the system of categories

involved, used frequently enough to maintain a separate designation outside the "other" category, and not properly placed into an already existing category.

For the majority of alternative categories, results revealed significantly higher correct coding rates or no differences, which, in turn led to more reliable and accurate labor force classification for the responses included in this study. When labels were changed to clarify category membership as well as when new separate and distinct categories were provided to accommodate responses previously difficult to code there was significantly better coding than when no category was furnished. Hence, the new and alternative categories helped to correctly code responses that were previously placed into a variety of different present categories.

These explicit alternative categories allowed more consistent and precise coding of responses since they were constructed to be exhaustive and appropriate for the type of responses actually given to the specific question. Consequently, these categories decreased the amount of interpretation that occurs when given difficult or ambiguous responses and, therefore, allow the FR to determine and code rapidly the category to which a response belongs.

In addition to the alternative categories' effect on coding error, the use of the alternatives is crucial for a number of additional reasons. The alternatives promoted a reduction in the use of the "other" category. Limited use of the "other" category aids to decrease both interviewer and respondent burden since each "other" response must be transcribed, thus, increasing the duration of the interview. Furthermore, as the categories were adapted from actual data, the use of these alternatives increased the amount of economic detail about the respondents that was previously not readily available because it was hidden in the "other" category.

In some instances, this study indicated that the present categories produced higher correct coding rates. One reason for this occurrence may be the lack of instructions on the usage as well as the definitions for the alternatives. It would be expected that appropriate training and the use of uniform established probes would improve coding of responses with these new categories. Further investigation of the use of alternatives should be conducted following training of FRs on their use.

There was a wide range in the rate of correct coding across categories, with some categories having correct coding rates under 50%. This may be an artifact of the data (i.e., due to the responses provided and the inability of the FRs to probe further, clarifying information). Moreover, it may suggest the need for additional refinement of the category choices. Finally, it may indicate the demand for improved training of FRs on the categorization of field coded open-ended questions.

Although both the coding and classification error was reduced, these results can only be applied to the responses which were selected for this study, as they did not represent the actual frequency of occurrence or the entire universe of

responses for each particular question. However, the responses chosen for the study were actual responses transcribed from the open-ended CPS labor force questions. Consequently, while the level of error may be different for the entire universe of responses, the direction would be expected to remain the same.

This research allowed critical observation of response type, the data, that are presently collected. The review of these data clarified the use, misuse, and nonuse of current categories as well as the ability and accuracy of coding both typical and unusual responses. Moreover, examination of the actual data and category use facilitated insight into the changing economic situation. For example, two categories--"plant-machine repair" and "material shortage"--were used infrequently. Since the 1940's when the CPS was designed, the manufacturing sector of the economy has become less prominent and the service sector has correspondingly become more important. Consequently, these categories were less appropriate for an economy based on more services and not used. In addition, the inspection of specific category usage illustrates the present categories' potential for coding error. Modifications could not have been suggested without an examination of response types typically coded as "other" or consistently coded into an inaccurate response category. For example, the "other" category for the question on absenteeism was overwhelmed with "maternity leave" responses. Consequently, review of the responses resulted in category modifications, new labels, additions, and deletions.

Several recommendations have ensued from this research. Findings suggest that separate and distinct categories, which allow the incorporation of response types given and provide sufficient description or definition of category membership, are needed for more reliable coding. Since the alternative categories seem to elicit more reliable coding with the responses employed, it is proposed that these categories be used for a larger field study to determine their efficacy. Moreover, results indicate that a review of the current data provided essential information on accuracy of coding. Hence, to alleviate further coding error, periodic evaluation of the data should be conducted.

Finally and most importantly, this study introduced a method for determining the reliability of response coding for field coded open-ended questions. Although this study is confined in its generalizability to the CPS because of the small set of responses employed, the usefulness of this method in studying field coded open-ended questions, in general, can be limitless.

REFERENCES

- Lazarsfeld, P.F. & Barton, A.H. (1955). Some general principles of questionnaire classification. In P.F. Lazarsfeld & M. Rosenberg (eds.) The Language of Social Research. Glencoe, IL: The Free Press.
- Schumann, H. & Presser, S. (1981). Open versus Closed Questions. In H. Schumann & S. Presser (eds.) Questions and Answers in Attitude Surveys.