EVALUATING INTERVIEWER STYLE FOR COLLECTING INDUSTRY AND OCCUPATION INFORMATION

David Cantor, Westat, Inc.
James L. Esposito, Bureau of Labor Statistics

Key Words: Probing, Cognitive Methods, Coding

1. Introduction

The Current Population Survey (CPS) is one of the primary sources of industry and occupation (I/O) data for the general population. It is the only source of these data that can provide current labor force and employment status information for particular I/O groups. Similarly, it is one of the few micro-data files that offers analysts repeated cross-sectional estimates of I/O distributions over both very short (e.g., monthly) and relatively long (one or more years) periods of time.

The variability associated with survey-based measures of I/O has been found to be quite high. This variability is the cause of some concern among analysts of not only the CPS, but also to analysts/policymakers using survey-based measures of I/O (BLS-NCES, 1979). For example, Collins (1975) found that 17% of the CPS sample changed 3-digit industry categories between consecutive monthly interviews and 32% changed 3-digit occupations.[1] This rate of change was much higher than those indicated by a direct report of change provided by the respondent at an interview administered one month later (about 2.5% gross change for both industry and occupation). A detailed analysis of the verbatim responses given at the two interviews attributed about 70% of the changes in the 3-digit codes to reasons other than an "actual" change (e.g., wording or coder problems).

The purpose of the research reported below was to focus on I/O error related to interviewer style.[2] Specifically, this research had two goals:

1. To better understand how interviewers affect accuracy of the I/O data on the CPS; and

2. To suggest guidelines that can be used for training interviewers on the I/O items.

Much of the variance in the responses to the I/O items stem from problems interviewers have with probing and recording information necessary to classify responses to a 3-digit category (Gaertner, et al., 1989). The focus of this research was to better understand the scope and magnitude of these problems with the goal of developing better training materials for the interviewers.

As used in this paper, two specific interviewer behaviors related to the collection process will be discussed (Esposito, 1990): (1) *the recording of information* provided by the respondent in reply to specific questions posed by the interviewer, and (2) *probing* by the interviewer when inadequate information is provided by the respondent. These two aspects of the process can have substantial impacts on all stages of the collection of I/O data. For example, if the interviewer does not probe properly, the respondent may provide inadequate or erroneous information. Similarly, if the interviewer does not record the information properly, the coder will be unable to correctly classify the case.

2. Research Questions and Design

In order to address both of these aspects of interviewer behavior, individual sessions with nine Census coders were conducted. In these sessions the coder was instructed to:

- Listen to tape recordings of CPS[3] interviews containing the I/O questions;

- Write down information from the responses that he/she felt would be necessary to code the case; and

- Critique the interviewer's questioning strategy. For example, was there information that the interviewer did not ask for? What kinds of probes could have been inserted to improve the information provided by the R?

Individual sessions with the coders lasted approximately one hour. During this time, each coder was asked to listen to six to eight separate reports of I/O. Three forms were filled out by the coder. The first form was used to record information from the interview that the coder thought was necessary for classifying the case. The second and third forms were used to administer questions to the coder on the quality of the interview (e.g., effectiveness of probes).

Two sessions were conducted simultaneously by two different monitors. The monitors controlled the tape recorder, administered debriefing questions and answered any questions the coders had during the process. Two monitors were used to expedite data collection.

Every effort was made to have at least two coders listen to the same interview. This allowed for calculating measures of reliability across coders. Some effort was made to overlap households across the two monitors as well. This was less successful, because of time constraints.

Additional information collected during the original CPS interview was obtained from the CPS data-tape. This included:

1. Information written down by the CPS interviewer; and

2. The codes assigned immediately after the CPS interview took place. After the CPS interview, the data are sent to Census's data preparation facility in Jeffersonville, Indiana. The I/O data are coded at this facility within a short period of time.

For purposes of discussion below, the code assigned from listening to the tape will be referred to as TRC (Tape Record Code). The code assigned during the CPS interview will be referred to as PIC (Phase I Code).

3. Analysis and Results

To investigate issues related to interviewer style, two research questions were addressed:

A. What is the difference between the interviewer and the coder descriptions and classifications?

B. What comments did the coders have on the interviewer's probes and questioning style?

In this section, the results that directly address each of these questions are reported. Prior to discussing these results, a general description of the sample is provided.

3.1 Sample

A total of 41 interviews were available for administration to the coders. Of these 41, 33 were completed with at least one coder. These 33 were selected out of the 41 by running through the list of eligible interviews in sequential order and finishing as many as possible in the one day allotted for collecting this information. The sequence of the 41 interviews was not, as far as we can tell, a function of anything related to questionnaire version.

Of the 33 interviews, six were listened to by one coder, 20 were listened to by two coders, and 7 were listened to by three coders.

Missing data resulted from several sources. First, several of the tapes were difficult to hear or inaudible to particular coders. Individual coders differed in their ability to understand and/or hear particular sequences. When the coder could not hear the tape, the data were coded as missing. Second, there were three household ID's that could not be located on the CPS data file. Supplementary information could therefore not be retrieved for these interviews. We suspect this was because of transcription of the ID. And third, there was a single interview where the respondent refused to provide industry information.

With only 33 interviews, the distribution of industries and occupations does not span across all I/O possibilities. Nevertheless, it does provide some variation for a number of different I/O situations. As one indication of this, the overall referral rate for these codes was about what occurred during the CPS interviews as a whole (17%). This would seem to indicate that the content of the descriptions presented about the same difficulty as an average CPS interview.

3.2 Agreement between Coders and CPS Interviewers

In order to measure the agreement between the information heard by the I/O coder and written by the interviewer, two codes were compared: (1) the code assigned by the I/O coder after listening to the tape (TRC), and (2) the code assigned at the time of the CPS interview (PIC). This comparison should provide an overall measure of the loss of information that is attributable to transcription of information by the interviewer.

Comparison of Detailed Codes. There are several ways to measure the difference between the TRC and the PIC. One such comparison is by counting up the total number of differences between each TRC-PIC combination. This excludes data that are missing for the reasons described above. Out of the remaining 125 comparisons, there were 19 (15%) that disagreed (Table 1). Disagreement here refers to either: (a) the codes not matching at the 3-digit level (four cases), or (b) one of the codes being coded as a referral[4] case (15 cases). The distribution by industry and occupation is relatively even.

One way to judge this rate is to compare it to the verification rate derived from the CPS production coding. This rate is calculated by having two additional coders classify a sample of cases. The rate at which the CPS coder agrees with at least one of these additional coders constitutes the "verification rate." This rate of 95% is a measure of the overall reliability of coding for the I/O data. Compared to this standard, the rate of 85% is slightly smaller.

Including "referral" as a legitimate code over-estimates the rate of disagreement, since these represent the most difficult cases to code.[5] To some extent, coders have discretion in judging which cases to refer and which to try to code. It is probably the case, therefore, that if these codes were ultimately assigned a 3-digit code, as in normal production coding, many would eventually agree. This situation was simulated by:

- Deleting those cases where the TRC was "referred;" and

- Substituting the PIC code assigned by the referral coder during the CPS coding process.

When this is done, the rate of disagreement drops by 50%. Out of the 103 3-digit TRC-PIC comparisons that remain, 8% disagree with the PIC. Unlike the previous comparison, however, the rate differs significantly by industry and occupation. Only 1 of the 8 differences are for the former, while 7 are for the latter. This is consistent with what one would expect, since there are fewer industry codes than occupation codes.

A second way to judge these data is to compare the agreement among the TRC and PICs for the same interview. This controls, to some extent, for coding reliability. The extent to which the PIC is in the minority *vis a vis* what one would expect from the reliability rate of 95% is an indication of the effect of listening to the tape and collecting information that the interviewer did not write down.

This comparison is restricted to the I/O reports where at least two coders listened to the tape. In this case, the PIC is in the minority 9% of the time (5 out of 53 possible comparisons).[6] This implies a verification rate (i.e., 91%) that is only slightly smaller than the production verification rate (95%).[7]

Overall, therefore, these data do not indicate a large portion of error that can be directly attributable to the information transcribed by the I/O coder.

Comparison of Transcribed Information. The TRC and PIC seemed to produce approximately the same set of codes, at least within error associated with coding reliability. To further explore differences, a detailed comparison of the transcriptions and debriefing information was made for those 19 cases where the TRC and PIC differed. This comparison was completed by a Westat I/O coder who attempted to pinpoint if the difference in the TRC and PIC was due to what the I/O coder heard or simply to an individual decision by one of the coders. To the extent that it is due to what was heard, one can attribute the difference to problems with what the interviewer had originally transcribed when conducting the interview.

In this comparison, it is assumed that the transcription provided by the I/O coder adequately represents what the I/O coder thought was important to classify the case. If information important to classifying the case was not written down by the coder, this analysis will mistakenly attribute the difference to an individual decision by the coder, rather than to something in the interview itself.

Of the 19 differences, 12 can be attributed to differences in the two transcriptions. For example, in one case the occupation code differed between TRC (387) and PIC (468). The wording between the two reveals that the difference in line 25C:

"school day care assistant" for TRC
"after school day care assis[tant]" for PIC

lead to the discrepancy. The TRC wording implies a connection to a teaching position to an elementary school teacher. The PIC wording implies more of a specialty in day care.

The critical differences in the transcriptions are not always due to more information in the TRC. This is indicated by the fact that the average of the number of words in each of the two sets of descriptions are identical (13.3 words).[8] Subtle changes in wording or the addition of a single word can account for the difference, as illustrated above. This emphasizes the importance of the interviewer writing down exactly what the respondent reports, with as much of the original detail as possible.

3.3 Comments by Coders on Interviewing Style

When listening to tapes, coders were asked to comment on the interviewers' probes. The purpose of these comments were to provide an indication of problems interviewers currently have with collecting information necessary to classify the job.

Coders were asked to provide both positive and negative comments. Examining these data for the 33 interviews played to the coders revealed that 23 of the 33 had at least one comment on either industry or occupation. These data indicate, therefore, that the coders commented at least once for about 70% of the interviews. Disaggregating by the number of positive vs. negative comments, the former outnumbered the latter by slightly less than 2 to 1 (20 vs. 11).[9] This last result is somewhat surprising. It was originally anticipated that the coders would primarily provide negative comments on the probes and interviewing style for these items.

The absence of many negative comments may have simply been a function of the small and selective nature of the sample of interviews. We suspect, however, it is also related to the difference between the coder's desire to easily classify the case and a concern with trying to come up with the "correct" code. This difference is related to the quality control procedures used by the Census when assigning I/O codes. These procedures consist of having a sample of cases classified by three coders. One of these coders classifies the case as part of the normal production of the CPS. The other two coders classify the case after the production coding is completed. To monitor quality, the three codes are compared. Coders are penalized whenever a code disagrees with the majority. There is no check on whether the majority code is actually "correct."

This process conditions the coders to classify the case into a category that is most likely to agree with what other coders may assign. Conflicting

information (e.g., between the job name and major activities) makes the coder's job more difficult, since it leaves doubt as to how another coder would classify the case. For this reason, the coders would like to discourage the interviewers from probing too much. The more information provided by the respondent, the greater the chance that there will be conflicting information. When listening to the tape, therefore, coders were usually content not to comment, as long as the information provided was enough to classify the case under existing rules. There was a clear tendency for the coders to discourage probing any responses that may have been ambiguous with regard to what the respondent was saying, but still "codeable."

Notwithstanding this last observation, these data provide some indication that much of the variability observed in the I/O data cannot easily be corrected by trying to improve interviewer's spontaneous, or "unstructured," probing skills.[10] If the coders cannot provide useful probes, then it would seem unrealistic to expect interviewers to do so. The interviewers are not familiar with the classification system and the task of generating probes is, apparently, quite difficult even for those familiar with the classification system.

At the very least, these results indicate that it would require an extremely skillful interviewer, well acquainted with the I/O system, to elicit both reliable and valid information. Substantially reducing error variability in the I/O has to come from - introducing structured questions that are consistent with how respondents think about and describe their jobs.

Both general comments and the specific ones related to the audio tapes can be divided into two groups. The first consist of comments where the interviewer should have tried to collect more specific information. Examples include:

- Did not ask whether the company was wholesale, retail or something else;
- Did not probe for specific information on the type of industry;
- Information provided on occupation type (the "C" line) was not sufficient for coding purposes. The response either referred to a general type of occupation or to activities within the occupation; and
- The major activities were not specific enough.

The second comment refers to situations where the information provided by the respondent is missing something that is critical to unambiguously classifying the case. In most of these cases, however, the interviewer is not in a position to know what information is required. For example, in one instance, the respondent described an industry as "kitchen ranges - manufacturing." In order to code the case, however, information on whether the ranges are commercial or household was needed.

With regard to the third comment, several interviewers said that inadequate occupation names pose a major problem. This frustration stems from the fact that the coding rules give this line the highest priority. Coders are instructed to try to code the occupation from this line and use the information on the activities line when necessary. If no clear occupation name is present, however, it is difficult for the coder to find a proper code. In response to this frustration, several coders suggested that interviewers be instructed to ask the respondent for an occupational title. This would provide the coder with a name that could be located among the eligible codes.

The fourth comment refers to descriptions of activities that do not provide the coder with adequate information to code a case. When this comment was made by a coder, it typically consisted of asking for more specific information about the content of the job (similar to the second comment above on industry).

The second type of problem relates to the interviewer asking for too much information. Examples include:

- Conflicting information on the major activity line was not resolved.
- The interviewer asked too many questions. She did not stop when there was enough information to code the case.

An extreme example of the coders objection to too much information is illustrated in their general comments that claimed that asking for major activities hurts as much as it assists in the coding process. These coders thought that since most of the information required to code a case comes from the name of the occupation, interviewers should concentrate on getting good information here. The activity line commonly conflicts with the information on occupation type and makes it more difficult to code the case.

4. Discussion and Recommendations

In this section, we discuss the results of our research relative to the two goals stated in the first section of the paper (i.e., to better understand how interviewers affect the accuracy of I/O data collected in the CPS, and to suggest guidelines that can be used for training interviewers on the I/O items).

4.1 Training Interviewers to Ask I/O Questions

The vast majority of the data-quality and coding problems mentioned by the coders during this research can be traced to the interviewer's lack of familiarity with the coding rules used by coders

and with the inherent complexities of the Census I/O classification system. If interviewers were aware of the coding rules, perhaps they could eliminate some of these ambiguities in descriptions. Five specific recommendations were made to accomplish this.

R1. Train interviewers to filter as little of the response as possible.

- *Record as much information as possible (within reason).* This includes the exact words used and, when possible, the order the words are spoken. Techniques that might assist the interviewer are:
  - -- Be ready to type in the answer as soon as the respondent speaks, and
  - -- To slow the respondent down when he/she is going too fast.

Given the fact that, on average, interviewers and coders recorded approximately 13 words in response to the two occupation questions, this does not appear to be an unreasonable task.

- *Repeat back what is written down for each question.* This includes verifying the spelling of names or technical terms used in the description.

This recommendation has to be balanced against several competing constraints. First, interviewers have very little space in which to write. Second, the use of dependent interviewing places a demand on constructing coherent descriptions that can be easily read by the interviewer at future interviews. Both of these constraints require the interviewer to filter information when recording I/O data. Nonetheless, both of the above procedural recommendations should improve the R's ability to capture critical information as the respondent describes his/her job.

R2. Train interviewers on the importance of obtaining an occupation name.

Probes that may be useful for this purpose are:

- "What do you call your job?"
- "What is the name of your occupation?"
- (proxy) "What does (he/she) generally call (his/her) job?"

R3. Give interviewers probes to use when listing multiple activities.

- In the last six months, which of these activities did you do most of the time?
- Which are your primary activities?

R4. Train interviewers as coders.

A more effective, but time consuming, way to acquaint interviewers with the coding rules would be to train them as coders. This was actually suggested by several I/O coders during this research. The training would not go into the more detailed rules I/O coders have to learn (e.g., when to refer a case, verification procedures, using auxiliary lists). It would simply provide them with the basic procedures coders go through and then have them classify a number of cases using these rules.

R5. Provide interviewers a reference document to assist in probing.

The primary concern here would be that it would make the interviewer's task more complicated. Looking up industry and occupation names while talking on the phone is not straightforward.[11] If a list is constructed, it would have to be relatively short and easy to use.

4.2 Understanding Interviewer Error

While this research was based on a relatively small number of cases, it does provide additional insight into problems related to collecting I/O information. At the outset, it was originally hypothesized that coders would have a number of comments on relevant probes that should be used. This did not turn out to be the case. In fact, there were more positive remarks about the interviewing style than negative. In support of this qualitative observation was the fact that a large majority of the codes assigned by listening to the tapes were identical to those assigned at the time of the CPS interview.

The small number of comments by the coders reflects, in part, the difference between coding reliability and validity. This high agreement between coders and interviewers, however, also reflects the fact that generating the correct probe is a very difficult cognitive task. The I/O system is very complicated and even though the coders are familiar with the system, supplying the correct probe requires complex comparisons between different types of codes. The implication is that it is probably asking too much of interviewers to generate probes without providing more structure in the questionnaire. As the questionnaire becomes automated, it is possible to accommodate this structure by developing more complicated branching systems. This system, for example, may use key words to provide structured probes to the interviewer.

References

BLS-NCES (1979). Feasibility Study For a Project on Improvement of Occupational Information. National Occupational Information Coordinating Committee. Performed under Contract Number AO-A01-788-00-1349 in the U.S. Department of Commerce. Published jointly by the Bureau of Labor Statistics and the National Center for Education Statistics.

Cantor, D and K. Levin (1992). **Summary of Activities to Evaluate the Dependent Interviewing Procedure on the Current Population Survey.** Final Report submitted to the Bureau of Labor Statistics, Contract No. J-9-j-8-0083.

Collins, C. (1975). "Comparison of Month-to-Month Changes in Industry and Occupation Codes with Respondent's Report of Change: CPS Mobility Study." Response Research Staff Report 75-5, May 15, 1975, U.S. Bureau of Census.

Esposito, Jim (1990) "Resolving the Gross Flow Problem (Research on the I/O Series)." Memo dated December 7, 1990 to Alan Tupek, Mike Welch and David Cantor.

Gaertner, G.H. Cantor, D. and N.L. Gay. **Research on Industry and Occupation Questions In the Current Population Survey.** Final Report submitted to the Bureau of Labor Statistics, Contract No. J-9-j-8-0083. June, 1989.

Rothgeb, J., Polivka, A., Creighton, K. and S.R. Cohany (1991). "Development of the Proposed Revised Current Population Survey" pp. 56-65 in **1991 Proceedings of the Section on Survey Research Methods,** American Statistical Association, Alexandria, VA.

Footnotes

1 This research was conducted as part of a larger research project to assist in the re-design of the CPS. (Cantor and Levin, 1992.)

2 Changes in the major I/O groups were about half these rates.

3 The CPS interviews were conducted as part of Phase I of the CPS redesign (Rothgeb, et al., 1991). Three different versions of the CPS were administered as part of this test. All three were used in the analysis discussed below.

4 Referral cases are those interviews where the coder could not classify the case. These cases are then "referred" to a senior coder when arriving at a final code.

5 The verification rate for codes referred is around 85%.

6 There was one case where the PIC agreed with one of the 3 TRCs assigned to the interview. This was not included in the 5 minority codes.

7 A second way to view this is by comparing all TRC's that were coded for the same case. Out of a total 104 such comparisons, there were 12 disagreements (11%).

8 When there was a refusal to answer the question or there simply was not any information, nothing was counted as a word.

9 Of the 23 interviews that had at least one comment, 8 had both positive and negative comments.

10 This is to be distinguished from structured probes that are written into the questionnaire.

11 This was evident from our experience trying to train I/O coders to classify respondents while talking to them over the phone.

Table 1. Number of Codes Assigned from the Tape Recording by Industry/Occupation and Agreement with Codes Assigned from the Phase I Interview*

	Agreed with Phase I Code?		
	No	Yes	Total
Industry	8 (13%)	54 (87%)	62 (100%)
Occupation	11 (17%)	52 (83%)	63 (100%)
Total	19 (15%)	106 (85%)	125 (100%)

* The tape recording code disagreed with the Phase I code when either: (1) one was referred and the other was not, or (2) the tape recording code differed from the non-referred Phase 1 code.