TESTS OF ALTERNATIVE QUESTIONS FOR MEASURING INDUSTRY AND OCCUPATION IN THE CPS

Gregory Gaertner, David Cantor, and Nancy Gay, Westat
and Susan Shank, Bureau of Labor Statistics
Gregory Gaertner, 1650 Research Blvd., Rockville, MD 20850

KEY WORDS: Response error, CPS, occupation, industry.

Industry and occupation (I&O) codes assigned to the same person in the Current Population Survey (CPS) typically show substantial month-to-month variation. This paper reports on research supported by the Bureau of Labor Statistics (BLS) to identify sources of response error in CPS I&O data and to design and test alternative questions to elicit more accurate I&O information. Data were collected from interviews conducted in a cognitive laboratory framework. Interviews were conducted with paid volunteers, followed by extended debriefings to validate interview information. The experimental alternative versions are compared with each other and with the current CPS questions in terms of measures of inter-coder reliability, agreement between employer and self-report, respondent and proxy reporting, and reliability over time.

INTRODUCTION

The CPS is one of the primary sources of I&O data for the general population. It is the only source that can provide current I&O information for particular demographic groups. Similarly, it is one of the few sources 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. Survey-based measures of I&O information such as those in the CPS are typically collected by asking a respondents several questions on the kind of work they do and the kind of businesses they work for. The verbatim responses to these questions are then converted by trained coders into three-digit I&O codes. The CPS, as most major surveys, utilizes the **1980 Alphabetical Index of Occupations and Industries** as the source of coding categories and information. The **Index** currently includes 231 industry groups and 503 occupation categories.

The variability of I&O information for the same individual over time has been found to be quite high. This conclusion is based on studies that have either re-interviewed respondents at two different points in time or used a separate measure of I&O to externally check the survey responses (Walsh and Buckholdt, 1970; Mellow and Sider, 1983; Collins, 1975). 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 detailed I&O (BLS-NCES, 1979). Most relevant for the CPS, Collins (1975) found that 17% of the CPS sample changed three-digit industry categories between consecutive monthly interviews, while 32% changed three-digit occupations.[2] These change rates were much higher than those indicated by a direct report of change provided by the respondent at the second interview (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 three-digit codes to reasons other than an "actual" change (e.g., wording or coder problems).

The task of collecting I&O information is a complex joint product. It requires that the interviewer help the respondent to order (or reorder) his/her job experience in terms of categories even the interviewer understands only imperfectly, and to communicate to the coder only the relevant information required for coding. As, the research cited above suggests, this joint product often leads to error.

Improvements to the questions asked are a reasonable starting point for improving the quality of I&O information. The I&O questions and associated probes are the tools common to the coder's, interviewer's and respondent's tasks. In this research, two types of changes were tested and evaluated. First, changes in the question wordings were tried to give more guidance to respondents in terms of referents, language and level of detail required. Second, context-specific probes were added to elicit the particular information required to distinguish among similar categories of occupations or industries; depending on the categories, this might be product produced, customer served, or skills required. Based on these general guidelines, three alternatives to the current CPS questions were developed for experimental testing.

THE CPS-1 AND TESTED EXPERIMENTAL ALTERNATIVES

Three alternatives to the current CPS I&O questions were tested, labelled ALTWORD, the NCT, and 'Type' Probe. Each was designed to improve both industry and occupation information. The ALTWORD version represented the least radical change, focusing on changes in question wording. The NCT alternative focused on particular hard-to-code occupations identified in the National Content Test.[3] For selected occupations, additional context-specific probes were added to the CPS-1 to obtain more detailed information on hard-to-code occupations. The 'Type' Probe alternative tested a probing strategy based on asking "What type of {...} was ...?" as a probe following the initial occupation question on kind of work done by the respondent. All of the alternative versions tested respondent self-classification into major industry group as a way of improving industry information.

Figure 1 summarizes the differences between the experimental alternatives and the current CPS I&O questions. Question-by-question differences between the alternatives and the current CPS questions are available from the contact author.

METHODS

The research tested these alternatives using the current CPS-1 questions as a control with 129 test subjects. Subjects were typically pairs of adults from the same household, interviewed concurrently. Subjects were debriefed extensively on the interviews to determine the nature and source of errors.

Subjects. Subjects were paid volunteers, interviewed in a cognitive laboratory setting. Interviews were conducted with 129 people overall; 48 pairs with both respondents working, five pairs with one partner working, and 23 singles. Subjects in the interviews were assigned systematically to different questionnaire versions and both members of a pair received the same experimental version or the CPS-1 as a control.

Interview Protocol. The interview protocol consisted of selected questions from the CPS control card, the CPS-1 interview (or alternative) for the subject and for his/her partner and debriefing on the interview. Respondents in pairs were also shown the responses their partners gave regarding their (the respondent's) job, and were asked whether they agreed with that information. Interviews were conducted in different rooms so that respondents were unaware of each others' responses. During the interview, permission was requested to recontact respondents to re-ask the I&O questions and to confirm the I&O codes assigned to them on the basis of the original interview. All interviews and debriefings were audiotaped for subsequent analysis.

Coding Procedures. Following the interviews, information from the I&O questions and the control cards

was reproduced and shipped to the CPS coding facility for coding. In order to produce measures of coding reliability similar to those used in the Census verification coding procedures, each case was coded by three independent coders. If the coder wished to refer a case, a referralist coder assigned the eventual code.[4]

Re-interviews. For paired respondents, one of the pair was selected for re-interview in order to minimize household response burden. When possible, we selected the member of the pair for whom coders disagreed in coding the original interview. If this person was not available, but the second member of the pair was, then he/she was interviewed. We were able to complete interviews with 70 of the 76 (92%) of the selected re-interview respondents.[5]

Employer Interviews. With permission of the respondents, their employers were interviewed by telephone two to three weeks after the original interview. Of the 124 respondents eligible for employer interviews, we secured a total of 69 permission forms. About two-thirds of the 55 respondents not participating refused to give permission; the remaining third were either self-employed or in some other situation where employer interviews were not feasible. We were able to complete interviews with 66 of the 69 respondents from whom permission was secured.

Measures of Validity and Reliability. Multiple indicators of I&O reliability and validity were collected. For self-responses, these measures included two basic measures of coding reliability:

1. **Amount of agreement among independent coders.** A major problem with the current series of questions is vagueness or ambiguity of responses. Interviewers and respondents may not provide the information required to unambiguously classify occupation or industry. An indicator of this for the study was whether three independent coders agreed on the same classification. Presumably, if information provided is adequate and relevant, disagreements among coders should be reduced.
2. **Rate of referral.** Coders tend to refer cases where information is ambiguous, inconsistent or incomplete. Thus referral rates should be less if questions are yielding more complete, consistent information. Further, since referral tends to be expensive, referral rate is indirectly a measure of coding efficiency.

Coder-based measures, however, do not directly address the issue of whether the "correct" industry or occupation was assigned. In order to directly address this issue, two other external measures of validity were collected:

3. **Employer interviews.** The responses collected from employers represent information from another source likely to be knowledgeable about the respondent's job. To the extent that agreement between employer and self-response is heightened, the external validity of the coding is increased.
4. **Respondent re-interview.** Re-interviews have been a common method used to measure the amount of error in I&O data (Collins, 1975). The assumption is that large changes in I&O between interviews is due to problems of measurement.

For proxy respondents, a similar set of coder-based reliability measures were collected. All proxy responses were classified by three independent coders and each coder was asked to fill out a confidence rating for each response. In addition, the paired design allowed for two additional comparisons involving the self and proxy report for the same subject:

1. **An evaluation by the subject of the verbatim proxy report.** During the debriefing, the self respondent was asked to evaluate the proxy report provided by the other member of the household.
2. **Direct comparison of coded responses.** Once the proxy reports were completed, the majority code was compared to the majority code derived from the self report.

RESULTS

This section compares the CPS-1 with the experimental questionnaire versions tested in terms of the measures of validity and reliability introduced above.

Results of the Comparisons -- Industry

Since the NCT and Type probe questionnaire versions were identical in their treatments of the industry questions, we combine them in this section. In the tables, we have added the supplementary label "major group self-classification" to describe this combined group. Likewise, we have added the label "goods/services specified" to describe the ALTWORD version more fully.

Table 1 presents the results of comparisons among the versions on coding performance and agreement between sources.

Coding Performance. The first column of Table 1 (% Referred) indicates the proportions of cases where at least one of the three coders coding the case referred it to a referralist coder. Overall 38% of cases were referred by one coder for self responses, and an identical 38% for proxy responses. The "% three-way agreement" in the second and third columns represent the proportions of cases that resulted in all three coders agreeing at either the three-digit (column 2) or major group (column 3) level.[6] The rates of agreement for industry are 83% at the three-digit and 87% at the major group level.

The rates of agreement indicate a fairly high probability that coders will assign a different code for the same description. At least one out of three coders assigned a different three-digit code 17% of the time. Second, note that the disagreement rates do not differ dramatically between three-digit and major group levels. This indicates that most of the disagreement crosses the major group categories. Disagreement is not simply assigning codes to closely related categories.[7]

While there are no statistically significant differences among the versions in terms of referral rate or rates of three-way agreement with this sample size, there is a non-significant tendency for CPS-1 to be referred more frequently than the experimental versions (50% versus 28% for the ALTWORD version, and 35% for the NCT/type probe versions).

Agreement Between Interview and Re-interview, Between Self and Proxy and Between Self and Employer Interviews. Additional comparisons among the versions can be made using measures of consistency between lab interview and re-interviews, between self and proxy responses, and self and employer interviews.

Seventy-eight % of industries coded from the re-interview matched the codes from the original interview at the three-digit level and 82% matched at the major group level (equivalent to Collins, 1975). If we consider the original interview to be the first month CPS interview and the second interview to be the second month, these figures would correspond to an 18% gross shift in occupation and industry at the major group level. These differences cannot be attributed to either actual changes in jobs (no respondents changed jobs in the interval) or to changes between self- and proxy-status across interviews.

Interview and re-interview responses to the experimental versions matched significantly more frequently than the CPS-1 at the three-digit and major group levels. Matches were found at the major group level using the CPS-1 less than 60% of the time, compared with more than 90% matches for the experimental versions. This suggests that the experimental versions tested might potentially reduce excessive month-to-month instability in industry estimates.

A majority of the employer reports of industry and occupation agree with the corresponding reports from self respondents. A relatively large minority, however, disagree. For industry, 85% of the cases agreed at the major group level, while 77% agreed at the three-digit level. There are no significant differences by version in match between employer and self-interviews although in both cases the results were consistent with the hypothesis that the experimental versions yield better information.

Summary of Comparisons -- Industry. We tested two variants of the CPS-1 version of industry questions. In all the experimental versions, respondents were asked to specify whether the kind of business the organization they worked for was mainly manufacturing, retail trade, wholesale trade or something else. In addition, in the ALTWORD variant, respondents were asked what product or service was produced/provided by the organization.

The results suggest that both of these variants had positive effects. Referral rates were lower for all experimental versions, and matches between interview and re-interview were more frequent. Other differences, while falling short of statistical significance tended to support the experimental versions as well. By implication, implementing the major industry self-classification question might reduce referral rates and improve coding accuracy. Further, to the extent that match between interview and re-interview approximates matches between CPS interviews, there is some reason to believe that self-classification might stabilize gross flow data as well.

The results also seem to support the use of the question asking for the nature of goods or services produced although differences in coder confidence, self-employer and self-proxy matches fell short of significance.

Results of the Comparisons -- Occupation

The differences among experimental versions are examined in Table 2 presenting coding performance and agreement among sources.

Coding Performance. Overall, 41% of the occupation self-responses were referred by at least one of the three coders responsible for the case. This proportion does not vary significantly by version; observed rates of referral vary between 38% for the CPS-1 to 44% for the NCT. Rates of three-way agreement likewise do not vary significantly by version. If anything, however, the results point to a superiority for the CPS-1. The average three-way agreement was 77% for the self responses, with versions ranging from 71% for ALTWORD up to 85% for the CPS-1.

Consistency Between Interviews and Between Respondents. There are no significant differences between versions in rates of interview/re-interview match, match between self and proxy responses and self agreement with coded responses.

Use of the NCT List and the Type Probes. These results suggest that none of the experimental versions yields better information than the CPS-1. One hypothesis to explain this result is that the probes made available by the experimental versions were not used frequently enough to make much difference in coding outcomes.

To test this hypothesis, we reviewed self-responses and coded them as follows:

- For the NCT and Type probe versions, cases were coded to determine (1) whether the NCT/Type list probe was appropriate (that is, was the respondent in one of the occupations targeted for a special probe), and (2) if it was appropriate, whether it was used correctly.
- For the CPS-1 and the ALTWORD versions where the probe was not available, cases were coded as to whether the probe would have been appropriate (that is, was the respondent in one of the targeted occupations).

The results of this indicated that for the NCT and Type probe samples, 18 cases should have been subject to the use of the list probe. In 17 of those 18 cases the probe was actually used. Thus, the list probe was seldom used inappropriately.

The generally nonsignificant result for the list probe could also be due to the fact that only 38 occupations even qualified for its use. Combining all occupations together as the tests described above did, may be masking differences for these particular occupations. One test of this is to compare coding performance when the list probe was used (the NCT and Type probe versions) with coding performance when the list probe would have been appropriate but was unavailable (the CPS-1 and ALTWORD versions).[8] In 20 of the 56 CPS-1 and ALTWORD cases (31%), the list probe would have been appropriate if it had been available. Results, unfortunately, are inconclusive, owing to the small sample sizes. Eight of the 17 cases (47%) where the list probe was available and appropriately used were referred, compared with 11 of the 20 cases (55%) where the list probe would have been appropriate but was not available. However, 12 of the 17 cases employing the list probe wound up in three-way agreements (71%) compared with 16 of 20 where the list probe would have been appropriate, but was not available (80%).

Results for the 'type' probe argue against its implementation. Of the 26 respondents asked the Type probe version, in one case the probe was used incorrectly and in four cases the probe was not asked when it should have been. The probe was frequently awkward in practice, even when used correctly.[9]

Summary of Occupation Coding Results. Three experimental versions of the current CPS occupation questions were tested. One version incorporated changes in question wording, while a second proposed a general probing strategy and a third offered specific probes for hard-to-code occupations. While substantial error in all measures of validity and reliability exist in coding occupation, there were few significant differences by version, and the pattern of non-significant differences suggest, if anything, superior performance for the CPS-1.

DISCUSSION

Experimental tests of questions developed to reduce error in I&O information were compared with the current CPS I&O questions. The results suggested that improvements in the quality of industry information were possible, by allowing the respondent to categorize him/herself in terms of major industry groups, and by specifying the sorts of goods or services produced.

The results also suggest that only limited gains in occupation coding reliability and validity can be gained by changing the wording of the questions. While there is substantial apparent error in I&O reporting, rates of error did not generally vary between the experimental versions and the current CPS questions. The minor differences in question wording tested (e.g., "main" instead of "most important" job duties) were too subtle to have much effect. Rather, the major sources of error in responses to the occupation questions probably come from respondents being unable to understand what is being asked for and limited knowledge of what constitutes an accurate response for coding purposes.

The specific probes for hard-to-code occupations were probably effective to the extent that they gave respondents a better idea of what was being asked for. If this is so, one source of improvement in occupation information could be gained by expanding the list of occupations for which context-specific probes are available. In the current hard-copy environment, this might be accomplished by providing interviewers with booklets listing occupational responses and additional probes. In a CATI/CAPI environment, however,

this extended probing strategy would be far more easily and extensively implemented. Artificial intelligence techniques could aid the extension of clarifying probes as additional experience accumulates.

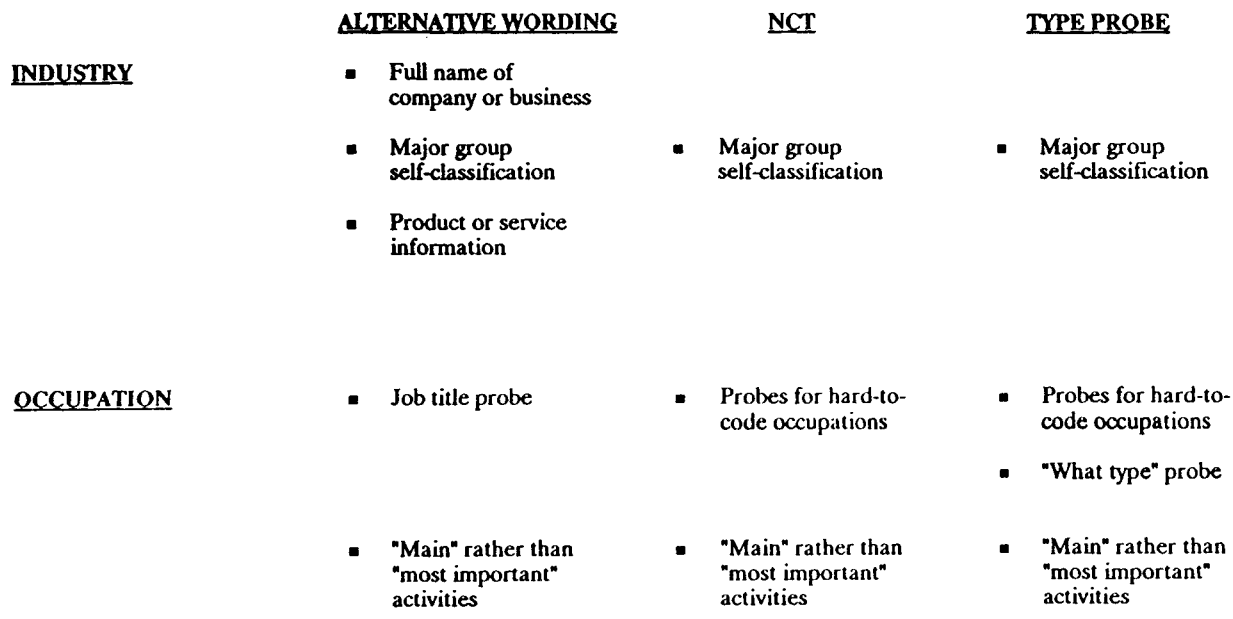

Figure 1 Summary of Differences Between Alternatives and CPS-1

Table 1

Comparisons of Tested Questionnaire Versions for Industry

		Three-Way Agreement			Lab Interview/ Reinterview Match			Self-Proxy Match			Lab Interview/Employer Interview Match		
Questionnaire Version	% Referred	3-Digit %	Major Group %	(N)	3-Digit %	Major Group %	(N)	3-Digit %	Major Group %	(N)	3-Digit %	Major Group %	(N)
CPS-1	50	82	85	(34)	57	57	(21)	67	70	(27)	72	78	(18)
Alt. Wording (Goods/services specified)	28	88	92	(25)	85	92	(13)	74	74	(19)	90	90	(10)
NCT/'Type' Probe (Major group self-classification)	35	83	87	(63)	85	85	(34)	69	78	(49)	76	84	(37)
TOTAL	38	84	88	(122)	76	79	(68)	69	75	(95)	77	83	(65)
$X^2(2)$	3.3	.4	0.6		6.3	9.3		.3	.5		1.2	.7	
p	.18	.80	.73		.04	.01		.88	.78		.54	.70	

Table 2

Comparisons of Tested Tested Questionnaire Versions for Occupation

Questionnaire Version	% Referred	Three-Way Agreement 3-Digit %	Three-Way Agreement Major Group %	Three-Way Agreement (N)	Self-Response/ Reinterview Match 3-Digit %	Self-Response/ Reinterview Match Major Group %	Self-Response/ Reinterview Match (N)	Respondent Agreement with Code %	Respondent Agreement with Code (N)	Self/Proxy Match 3-Digit %	Self/Proxy Match Major Group %	Self/Proxy Match (N)
CPS-1	38	85	88	(34)	81	90	(21)	62	(21)	65	77	(26)
Alt. Wording (Goods/services specified)	38	71	75	(24)	69	77	(13)	64	(11)	56	67	(18)
NCT	42	73	88	(33)	67	78	(18)	47	(17)	68	72	(25)
'Type' Probe	43	77	80	(30)	69	75	(16)	43	(14)	71	75	(24)
TOTAL	41	77	83	(121)	72	81	(68)	54	(63)	66	74	(93)
$X^2(2)$	.3	2.2	2.5		1.2	1.9		2.0		1.2	.6	
p	.96	.54	.47		.74	.60		.58		.76	.89	

FOOTNOTES

[1]Changes in the major I&O groups were about half of these rates.

[2]See Hodge and Siegel (1966) for other more detailed examples.

[3]The National Content Test was a test of new occupation questions for the 1980 Census. A description of the test can be found in BLS-NCES, 1979. Some discussion of results can be found in memos by Priebe (1977a, 1977b).

[4]This was accomplished by arbitrarily dividing paired respondents into two groups, 'A' and 'B'. Then, three coders coded the A group and three different coders coded the B group. This ensured not only that each case would be coded by 3 independent coders, but also that no coder would code both a case and a partner's proxy report, since the two should describe the same job. To ensure that no referralist would code the same case twice, each coder passed referrals to a different referralist.

[5]76 households were represented in the original interviews; 54 paired respondents plus 22 single respondents.

[6]This comparison is based on the final 3-digit code assigned. If a production coder referred a case, the code assigned by the referral coder was used.

[7]These rates of referral and three-way disagreement are much higher than are found in actual CPS coding. This likely reflects differences between the procedures used for the lab and those used in production verification. The laboratory study used questionnaire forms that the coders were not familiar with, had the coders classify responses during non-CPS weeks and were filled out by non-CPS interviewers. The results of this research, therefore, should be taken in light of the additional variability introduced because of these differences from normal CPS procedures.

[8]While the test is also appropriate for proxy responses, time and resource constraints stopped us from coding the proxy responses in terms of their appropriateness for the list probe.

[9]We cannot perform an analysis similar to that for the list probe since it was always technically appropriate to ask "What type of {...} is ...?" however awkward it might be in practice.

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.

BLS (1988). *BLS Handbook of Methods*, U.S. Department of Labor, Bureau of Labor Statistics, Bulletin 2285.

Bodenhausen, G., and Wyer, R.S. (1987). "Social Cognition and Social Reality: Information Acquisition and Use in the Laboratory and the Real World" pp. 5-41, in Hippler, H., Schwarz, N. and S. Sudman (eds) *Social Information Processing and Survey Methodology*. New York: Springer-Verlag.

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

Hauser, R., and Featherman, D. (1977). *The Process of Stratification: Trends and Analysis*. New York: Academic Press.

Mellow, W., and Sider, H. (1983). "Accuracy of Responses in Labor Market Surveys: Evidence and Implications." *Journal of Labor Economics*, 1: 331-344.

National Occupational Information Coordinating Committee (NOICC), "Feasibility Study for a Project on Improvement of Occupational Information" (November, 1979). National Research Council, National Academy of Sciences, Final Report, Contract AO-AO1-78-00-1369.

Priebe, J. (Jan. 7, 1977). "Inclusion of Occupation Job Title Question in the 1980 Census." Memo to Paula Schneider, Chief, Labor Force Statistics Branch.

Priebe, J. (Jan. 6, 1977). "NCT - Frequency Distributions for New Occupation Question with Class of Worker Comment." Memo to Paula Schneider, Chief, Labor Force Statistics Branch.

Taylor, D. (1976). "The Accuracy of Respondent-Coded Occupation." *Public Opinion Quarterly*, 40: 245-255.

U.S. Census (1978). *An Evaluation of 1970 Census Occupational Classification*, Technical Paper No. 41, U.S. Census. Washington D.C.: U.S. Government Printing Office.

Walsh, T., and Buckholdt, P.J. (1970). *Accuracy of Retrospectively Reporting Work Status and Occupation Five Years Ago*. U.S. Bureau of the Census Working Paper PA-(75).

Winch, R., Mueller, S., and Godiksen, L. (1969). "The Reliability of Respondent-Coded Occupational Prestige." *American Sociological Review*, 34: 245-251.