RECORDED INTERVIEW BEHAVIOR CODING STUDY
NATIONAL SURVEY OF RECENT COLLEGE GRADUATES

Margaret Cahalan, Susan Mitchell, Mathematica Policy Research (MPR);
Lucinda Gray, Selma Chen, Westat; John Tsapogas, National Science Foundation

KEY WORDS: Behavior coding, questionnaire design, cognitive methods

1. INTRODUCTION

In behavior coding, aspects of interviewer and respondent behaviors are coded as a means of identifying questions that may need to be reworded or redesigned, or questions that are more difficult for respondents to answer. For these questions, the coding may also indicate areas where additional or different interviewer training is needed.

Behavior coding has been used in several studies to test the extent to which interviewers follow strict rules of behavior to ensure a standardized survey administration (Kahn and Cannell 1957; Okensberg, Cannell, Kalton, 1991). If the question is awkward or unclear, or the material is difficult, the respondent will frequently interrupt and interviewers may be forced to reword questions or provide extensive clarification. By coding such things as the number of times the interviewer deviates from the written script or has to probe, or the number of times the respondent asks for clarification, one can identify problem questions.

Other researchers have noted that there is a tension between the survey interview as an interaction event and as a neutral measurement instrument (Suchman and Jordan 1990). They note that, compared with ordinary conversation, the survey interview "suppresses those interaction resources that routinely mediate uncertainties of relevance and interpretation." They find that survey data are potentially undermined by the same prohibition against interaction that is intended to ensure reliability. They argue against mistaking sameness of words for stability of meaning. Thus, this approach concentrates on also trying to observe the extent to which standardized meaning was communicated rather than just the extent to which deviations from a written text occurred.

2. APPROACH

Drawing upon both approaches described above, we used behavior coding as a means of evaluating the reliability and validity of questions on the 1993 National Survey of Recent College Graduates (NSRCG). The NSRCG, sponsored by the National Science Foundation and conducted by Westat, is an on-going survey of bachelor's and master's degree recipients in science and engineering. The survey collects data on the educational backgrounds, employment outcomes, and demographic characteristics of recent degree recipients. In 1993, the sample comprised 25,785 graduates who received their degrees in academic years 1990-91 and 1991-92. Sample members were interviewed using computer-assisted telephone interviewing. The (unweighted) response rate was 86 percent.

For the behavior coding study, we examined interviewer compliance with interview scripts, interviewer interactions with respondents, and how well question meaning was conveyed. The objective was to improve the questionnaire for administration in 1995. The coding observed the extent to which the interviewers deviated from the structured questions. It also included the observation of whether, in the view of the behavior coder, the meaning of the question was adequately communicated and a "correct" response was obtained by the structured question. This "correctness" was determined primarily by observing the quality of responses to the question.

2.1 Coding Scheme

For the NSRCG behavior coding study, eight elements were coded for each question. The coding scheme was a modification of schemes used at Westat for other surveys (Shepherd and Vincent in 1991 for the National Postsecondary Student Aid Study; and Cantor, Levin and Carnes in 1993 for the National Household Education Survey). Four of the coded elements primarily concerned interviewer behavior and four related to respondents or their response. Each of these elements was coded for each item on the survey:

Interviewer Behavior

1. Was the question asked, that is, were the skip patterns followed correctly?
2. Was the question asked exactly as written?
3. Did any probing or clarification on the part of the interviewer occur and was it appropriate?
4. Did the interviewer demonstrate affect (laughter, sympathy) and if so was it appropriate?

Respondent Behavior

1. Was a "correct" response obtained (without a probe, with probe, keyed correctly)?
2. Was a repeat of the question requested?
3. Was a clarification requested?
4. Was there any indication on the part of the respondent that the question was sensitive?

2.2 Operational Methods

Live interviews with ninety-eight NSRCG respondents were taped during a period midway through the data collection period (August 7-8, 1993). If the respondent gave permission for the taping, the CATI interviewer conducted the interview in the standard fashion. Respondents were selected at random for the taping, with the only requirement being that they gave permission for the taping.

To prepare for behavior coding, responses to the CATI interview were copied onto blank questionnaires. These questionnaires allowed behavior coders to easily compare what they heard on the tape with what was recorded by the CATI interviewer. Each item on the ninety-eight questionnaires (a total of 11,581 items) was coded for interviewer and respondent behaviors.

3. FINDINGS

Table 1 contains the number and percent for each interviewer and respondent behavior code over all questions. As show in Box A, responses were autocoded by CATI 8 percent of the time and interviewers verified a response given ahead 2 percent of the time. Since this was a CATI study, there were almost no skip pattern errors (this could occur only if an item was keyed incorrectly).

The remaining interviewer behavior codes (Boxes B-D) and the respondent behavior codes (Boxes E-H) were assigned only for cases where the question was asked or verified by the interviewer (10,586 cases). Autocoded cases were excluded.

3.1 Interviewer Behavior

Variation from the text. Box B summarizes the amount of times that the interviewers read or did not read the question exactly as worded. As can be seen from the table, there was a very high degree of adhering to the exact wording of each question. In 97 percent of the 10,586 questions coded there was no variation recorded. In 2 percent of the questions coded (213 times) a minor variation that was judged by the coders as non-biasing occurred, and in 3 cases a variation

occurred that was judged biasing. In an additional 1 percent of the cases (152 times) the respondent answered ahead.

Probing and clarification. Probing or clarification is usually done when a response is not forthcoming or when the respondent gives a response that the interviewer thinks indicates that the question was not understood. For the NSRCG, interviewers probed or clarified in approximately 9 percent of the 10,586 questions coded. In almost all of these cases (932), the probes were judged by the behavior coders to be appropriate. In 9 cases the probe was judged to be wrong or not useful but not biasing, and in 9 additional cases it was judged to be biasing.

Affect. We also coded whether there was affect displayed by the interviewer in administering the questions. An example might be indicating that a specific field of study was a good one or expressing sympathy for someone who could not find a job. Laughter is another example of affect. Appropriate affect was coded as occurring about 1 percent of the time (93 cases). In 10 cases the coders found affect that was inappropriate or leading.

3.2 Respondent Behavior

Response. Coders were asked to make a judgment as to whether a "correct" response had been obtained from the interview. This required a decision as to whether the meaning of the question was conveyed in a manner sufficient to elicit the "correct" response. In the view of the coders, respondents gave answers that appeared correct 99 percent of the time (93 percent of the time without probing, and an additional 6 percent of the time after probing). Respondents gave qualified responses less than 0.5 percent of the time.

Repetition, clarification, sensitivity. Respondents requested repetition of a question 1 percent of the time and clarification 2 percent of the time. Clarification was done without being requested 4 percent of the time. Almost none of the respondents indicated that questions were sensitive (5 times), even though the questionnaire asks for annual salary and detailed demographic data.

While these responses indicate that overall the interview did not involve large amounts of unstructured (unwritten) behavior on the part of the interviewers, and that respondents typically answered without requesting clarification, examination of specific questionnaire items reveals that some questions had relatively larger amounts of unstructured behavior than others. The next section examines the data for selected items.

TABLE 1. SUMMARY BEHAVIOR CODES FOR ALL QUESTIONS

Interviewer Behavior (A-D)			Respondent Behavior (E-H)		
A. Asked	Number	Percent	E. Response	Number	Percent
Total	11,581	100%	Total	10,586	100%
Asked	10,398	90	Correct	9,858	93
Incorrectly skipped	13	*	Correct after probe	630	6
R answered ahead-verified	188	2	Qualified (expressed uncertainty)	47	*
R answered ahead-not verified	39	*	Clearly incorrect and recorded	1	*
Autocoded	935	8	Not obtained	32	*
Nonresponse	8	*	Keyed incorrectly	18	*
B. Variation from Text			F. Repeat		
Total	10,586	100%	Total	10,586	100%
No variation	10,218	97	Not requested	10,402	98
Minor non-biasing variation	213	2	Requested	143	1
Major biasing variation	3	*	Done not specifically		
Answered ahead-verified	152	1	requested	41	*
C. Probing/clarifying			G. Clarification		
Total	10,586	100%	Total	10,586	100%
Not done	9,636	91	Not requested	9,935	94
Done appropriately	932	9	Requested	231	2
Wrong probe	9	*	Done, not specifically		
Biased probe	9	*	requested	420	4
D. Affect			H. Sensitive		
Total	10,586	100%	Total	10,586	100%
Not done	10,483	99	Not indicated	10,581	100
Done appropriately	93	1	Indicated	5	*
Done unnecessarily and maybe leading	8	*			
Done inappropriately or leading	2	*			

*Less than .5 percent

3.3   Selected Questionnaire Items

In order to gain an indication for a given question whether any additional interaction occurred, other than simply asking the question without variation and recording a "correct" response, we combined the responses in such a way as to indicate whether any one of the other coded behaviors had occurred. Those cases that have no other behaviors are referred to as questions that were "asked and answered only."

Of the total 10,586 questions coded, 85 percent were "asked and answered only." Thirteen percent had at least one instance of non-scripted behavior; usually this was probing, which occurred most frequently.   In the   following   sections,   we   examine   selected questionnaire items that had less than the average of 85 percent "asked and answered only."   These are employment and education questions that are used on a number of NSF surveys and that had higher than average   incidence   of   interviewer-respondent interactions.

3.4 Employment Questions

The questions with less than 85 percent in the "asked and answered only" category appear in Table 2. The table shows the sample size, the percent asked and answered only (with no wording variation), and the percent with selected interviewer and respondent behavior codes.

Labor force status. This question asks, "Were you working for pay or profit during the week of April 15, 1993? Please include self employment and any paid jobs from which you were temporarily absent, for example, for illness, vacation, or parental leave." Interviewers had minor wording variations 11 percent of the time and probed or clarified 10 percent of the time. Respondents requested repetition 4 percent and clarification 4 percent of the time. Wording variations occurred most often when respondents answered the question before the interviewer completed the question, a common occurrence in long questions.

Reasons not working. This questions asks respondents who were not working to answer "yes" or "no" to a list of possible reasons why. Interviewers probed or clarified 40 percent of the time and displayed appropriate affecting behavior 20 percent of the time. Respondents requested clarification 7 percent of the time.

Employer information. All of the employer questions had higher than average rates of being probed or clarified, ranging from 9 percent for whether the employer was an educational institution to 23 percent for the employer name. Interviewers reported that respondents often questioned why they were being asked for employer name and city and how this information would be used.

Salary. Interviewers probed or clarified for salary at higher than average rates, with 38 percent for principal salary amount, and 31 percent for principal salary unit (annual, weekly, monthly, etc.). Requests for clarification were initiated by respondents at much higher than average rates, 22 percent for principal salary amount and 17 percent for principal salary unit.

Work activities. This question reads, "The next question is about your work activities on the principal job you held the week of April 15. For each of the following work activities, please tell me whether the activity occupied 10 percent or more of your time during a typical work week on this job. Please answer yes or no for each item." The first activity on the list had a slightly lower than average "asked and answered only" rate, 83 percent. This is probably due to difficulties with the question itself, rather than any problem with this particular activity ("accounting, finance, contracts"). Interviewers probed or clarified for 11 percent of the cases and had minor wording

variations for 6 percent. Respondents requested clarification 5 percent and repetition 1 percent of the time.

The question asking for the activity on which the respondent spent the most hours had a 25 percent probed or clarified rate, and respondents requested repetition 4 percent of the time. Interviewers reported that respondents sometimes had difficulty choosing between activities, or reported equal time spent on two or more.

Supervise others. This question asks, "Did you supervise the work of others as part of this job? Please answer yes if you assign duties to workers and recommend or initiate personnel actions such as hiring, firing, or promoting." Some respondents had difficulty with this instruction, as reflected in the 15 percent probed or clarified rate for this question. When asked for the number of people supervised, 11 percent of the cases required probing or clarification, and 7 percent of the respondents gave a qualified response about which they expressed uncertainty.

3.5 Education Questions

The following education questions with less than 85 percent in the "asked and answered only" category also appear in Table 2.

School name. Interviewers probed or clarified for college or university names at a higher than average rate (28 percent for the degree college and 23 percent for the college attended during the reference week, if any). Since interviewers were asked to record the complete school name, they may have probed to get the whole name or to obtain the specific campus.

Major field of study. Interviewers probed or clarified 17 percent of the time when asking for respondents= major field of study when they first enrolled, and 10 percent of the time when asking for the degree major. The degree major had 8 percent minor wording variation and was answered ahead 5 percent of the time. Interviewers were instructed to reference the degree information (type, date, or school name) as needed to ensure the graduate was answering for the correct degree. This encouraged some minor wording variations.

Loan amounts. Four items on loan amounts were included on the survey: amount borrowed for undergraduate degrees, amount owed for undergraduate degrees, amount borrowed for graduate degrees, and amount owed for graduate degrees. Because of small sample sizes and similarity of responses, the behavior codes for these items were combined. Interviewers probed or clarified 31 percent of the time, had minor wording variations 4 percent of the time, and displayed fecting behavior 3 percent of the time. Respondents

TABLE 2.   SELECTED QUESTIONS WITH LESS THAN 85 PERCENT "ASKED AND ANSWERED ONLY" RATE

Question	Sample Size	Asked And Answered Only	Interviewer Behavior				Respondent Behavior		
			Minor Wording Variation	Answered-Ahead-Verified	Probed or Clarified Appropriately	Appropriate Affect	Qualified Response	Repetition Requested	Clarification Requested
Labor Force Status									
Working during reference week	97	73%	11%	0%	10%	0%	0%	4%	4%
Reasons not working	15	47	7	7	40	20	0	0	7
Date last worked..	22	68	0	0	32	0	5	0	0
Employer Information									
Employer name	86	72%	2%	0%	23%	0%	0%	0%	2%
Educational institution?	44	57	14	21	9	2	0	5	0
Salary									
Amount...	86	41%	2%	0%	38%	4%	2%	2%	22%
Unit	84	50	1	1	31	2	0	2	17
Work Activities									
Work activity - Accounting	86	83%	6%	0%	11%	0%	0%	1%	5%
Work activity - most hours	68	54	1	0	25	0	2	4	0
Supervise Others									
Whether supervise others.............................	86	83%	2%	0%	15%	0%	0%	2%	2%
Number supervise directly	28	79	0	0	11	4	7	0	4
School Name									
College for degree....................................	40	65%	3%	0%	28%	3%	0%	0%	3%
College attended since degree, if any	26	77	0	0	23	0	0	0	4
Major Field of Study									
Major when first went to college...............	96	77%	0%	0%	17%	1%	0%	2%	4%
Degree major	40	80	8	5	10	0	0	0	0
Other Education									
Loan amounts...	113	57%	4%	0%	31%	3%	7%	4%	7%
Grade point average	96	77	0	0	16	1	3	1	6

requested clarification 7 percent of the time and repetition 4 percent of the time. In addition, respondents gave qualified responses about which they expressed uncertainty 7 percent of the time.

Grade point average. This question asked "Using a 4-point scale, what was your overall undergraduate grade point average (GPA)?" Answer categories were read only as needed. For this question, interviewers probed or clarified 16 percent of the time and respondents requested clarification 6 percent of the time.

4. CONCLUSIONS

Overall, the results of this analysis indicate that the large majority of questions about employment and education in the 1993 NSRCG were read as written by the interviewer and understood by the respondent. In 97 percent of the 10,586 questions coded, the interviewer read the questions exactly as scripted; in 9 percent of the cases, the interviewer probed or clarified. The questions with higher than average rates for problem codes were generally of the following type:

1. *Long questions.* These questions sometimes included a qualifying statement about what to include, for example, the question about employment status during the reference week, or the salary question.
2. *Questions asking about sensitive behaviors.* For example, the question about reasons why respondents were not working during the reference week had a higher than average probed and clarified rate, and interviewers were more likely to show affecting behavior.
3. *Introductory questions to items in a series.* Probing or clarifying generally was needed more frequently for these questions, probably stemming from confusion about the series itself rather than the individual question.
4. *Questions asking for detailed information.* These include, for example, employer name, occupation, and name of degree-granting institution. Since interviewers were asked to get complete answers, they may have probed to get the whole name, or in the case of institutions, the specific campus.
5. *Questions asking about information that may be difficult to recall.* Questions in this category include date last worked and number of people supervised. These questions had higher than average rates of respondents giving answers about which they expressed uncertainty.

The most frequent behavior code was used for interviewer probing or clarification of questions. In almost all of these cases (98 percent), the interviewer responded appropriately. This is because they were often instructed to probe or clarify in order to obtain complete answers that were spelled correctly.

These positive results, however, may also relate to the fact that interviewers knew they were being recorded and for what purpose. This probably influenced their behavior to conform to the questions as scripted, avoid affecting behavior, and to probe and clarify appropriately. Similarly, knowing they were being recorded may have influenced respondents to pay careful attention to the questions and request clarification less often. In actual practice, across a wider range of interviewers and respondents, problems would likely be more prevalent.

These results point out the type of questions that are subject to respondent and interviewer error, and suggest trade-offs to consider when formulating questions (for example, balancing the need for detailed information with expected higher rates of probing or clarifying and the increased time to administer). The results also indicate types of questions where enhanced interviewer instruction or training may be needed: long questions, introductory questions to items in a series, questions about sensitive behaviors, and questions asking for detailed information or information that may be difficult to recall. Researchers who ask questions about employment or education may find the results useful in formulating questions that yield reliable results. Further study is needed, however, to explore the exact nature of these problems and their consequences on data quality. For more precise estimates of the incidence of the interviewer and respondent behaviors noted in this report, we recommend the study be replicated across a larger group of interviewers who are unaware of the purpose for which their interviews are being recorded.

References

Kahn, R.L. & Cannell, C.F. (1957) *The Dynamics of Interviewing*, New York: John Wiley & Sons.

Oksenberg, L., Cannell, C., & Kalton, G. (1991) "New Strategies for Pretesting Survey Questions." *Journal of Official Statistics*. 7:349-365.

Suchman, L. & Jordan, B. (1990) "Interactional Troubles in Face-to-Face Interviews." *Journal of the American Statistical Association*. 85:232-91.