PROCEDURES FOR VALIDATING REPORTS OF HOURS WORKED AND FOR CLASSIFYING DISCREPANCIES BETWEEN QUESTIONNAIRE REPORTS AND VALIDATION TOTALS

W. Sherman Edwards, Westat, Inc.
Roger Levine, American Institutes for Research
Sharon Cohany, Bureau of Labor Statistics
Contact Sherman Edwards, Westat, 1650 Research Blvd, Rockville, MD 20850

KEY Words: CPS, labor force statistics, respondent debriefing

Introduction

In order to determine the effectiveness of alternative questionnaire wordings or survey administration procedures in eliciting recall of factual information, it is critical to possess "true" values for this information. Often, respondent reports can be validated independently of the respondent. However, in some situations, no independent source is any more reliable than (or even as reliable as) the self-respondent. Reports of hours worked during a reference week, when "hours worked" includes all hours spent on job-related activities, is an example. In such situations, potential "verification" information may be subject to as much or more error than that obtained from a self-respondent.

Once questionnaire reports have been validated, and the magnitude and direction of errors calculated, survey designers may still require further information to guide remedial design efforts. It would be desirable to know, for this purpose, why respondents made errors -- did they misunderstand the question, did they just not have the information being requested, did they not try hard enough to remember, was their recall or estimation process flawed or inappropriate, or did they "censor" the response for some reason? Depending upon the answers to these questions, quite different remedial design strategies might be appropriate.

This paper describes a protocol for validating reports of hours worked during the previous week and a system for classifying errors (discrepancies between questionnaire responses and "validating" responses). Both the validation protocol and the classification system were derived from recent work in applying the principles of cognitive science to survey design. The paper is based on research done on the Current Population Survey (CPS) "hours worked" series by Westat and the American Institutes for Research (AIR), under contract to the Bureau of Labor Statistics (BLS). The research, one of a series of studies to support redesign of the CPS, is described in Edwards, Levine, and Allen (1989). The two principal objectives of this research were (1) to evaluate proposed variations of the CPS hours worked question series and (2) to assess the differences in the accuracy of hours worked as reported by self-respondents and proxy respondents. The analysis focused on the discrepancies between reports of hours worked elicited by current and proposed question items and reports elicited through validation procedures from the subjects themselves. Limitations of the research are detailed in Edwards, Levine, and Allen (1989). They include the differences in interview environment between a "laboratory" and actual field settings, the use of non-Census interviewers, restriction to first-month CPS interviews, and non-random sample selection procedures.

The remainder of this introduction describes the CPS and the rationale for the current research. Section A below describes the "validation" protocol in detail and discusses its application in the determination of error magnitude and direction. Section B describes the classification of discrepancies obtained with this protocol into different types of response error.

Background -- The Current Population Survey

The Current Population Survey (CPS), the nation's primary source of information on employment and unemployment, is administered monthly by the U.S. Bureau of the Census for the U.S. Bureau of Labor Statistics to a probability sample of about 56,000 households. Households remain in the sample for a total of 8 months -- they are in for 4 consecutive months, leave the sample for 8 months, and then return for another 4 months. First- and fifth-month interviews are conducted in person, while most other interviews are done over the telephone. The respondent can be any responsible person in the household age 14 or older, and may change from month to month. The reference period for the survey is the calendar week that includes the 12th of the month; the survey is conducted the following week. The same interviewer usually conducts all interviews for a particular household.

As part of an overall redesign of the CPS for the 1990's, BLS commissioned Westat and the American Institutes for Research (AIR) to investigate the possible sources of error in the hours worked series, particularly question wording and order and the use of proxy responses. This task led to the development of the "validation" and coding protocols described in this paper.

A. The use of a debriefing protocol to "validate" responses

Introduction

Since the research objectives were to evaluate questionnaire alternatives and to assess the relative accuracy of self- and proxy responses to questions on hours worked, two design decisions were made quickly: the research would involve interviewing pairs of respondents from the same household concurrently, asking each about him/herself and the other person; and the questionnaire variations would be administered so far as possible as if the interviews were "real" CPS cases. These procedures would yield self- and proxy reports of hours worked for each subject, and would allow analysis of agreement between these two reports. However, to assess the accuracy of responses, it would be necessary to obtain "true" values for the number of hours worked in the previous week.

A number of possible validation schemes were considered, all of them flawed to a greater or lesser extent. Some options (such as having subjects maintain diaries before the interview) were discarded as potentially biasing. That is, the "validation" procedure would affect the responses to the survey questionnaire. Other options were themselves fraught with potential error. For example, obtaining a proxy report based upon memory or estimation from a third party, such as a subject's work supervisor, was viewed as no more likely to be accurate than the report of the subject or a spouse.

Two methods of validation were selected: a detailed self-report obtained after the CPS-like interview, and a record-check validation from the employers of subjects agreeing to the procedure. Neither of these methods was considered to be able to yield the full "truth" -- self-reports because there was no guarantee of perfect motivation, recall, and response formulation, and employer records because many of the kinds of work asked about in the CPS are not documented (the hours worked by self-employed persons, the work of many salaried employees, unpaid "extra" hours, and so on). Further, obtaining information from records is itself subject to "response error." (Marquis, 1984)

Methods

Seventy-one individuals served as subjects. These subjects were recruited purposively, to include individuals

with characteristics of particular interest to the research project -- that is, individuals with multiple jobs, part-time jobs, or many "off the book hours," as would be characteristic of some managers or other professionals. House Market Research (HMR), a private market research firm, recruited these individuals. Subjects were offered $15 incentive payments for their participation. In order to investigate proxy effects, 28 pairs of individuals from the same household were recruited. Five of these 56 persons did not work during the previous week. The remaining 15 subjects (all employed) came from 15 separate households.

Subjects traveled to HMR facilities, where their informed consent to participate in this research was elicited. Subjects were individually interviewed in separate rooms, so that paired respondents would not overhear each other's responses. The protocol began with administration of the CPS Control Card and one of four different questionnaire variations, simulating an actual CPS interview. The hours worked variations, identified by letter, comprised: (K) the CPS-1, amended to ask separately about hours worked at main and other jobs; (L) a variation developed by the CPS Questionnaire Design Task Force, asking first about usual hours worked, then exceptions during the reference week, and finally for hours actually worked during the reference week; (M) variation K with the addition of a "motivational" introductory statement and a calendar as recall aid; (N) variation K with detailed definitions of key terms read to the respondent.

An extensive debriefing followed administration of the simulated CPS interview. The debriefing protocol included the self-validation procedure, questions about the respondent's understanding of work-related concepts, the response task, and their recall strategies, and questions related to a concurrent research project on the industry and occupation questions. All interviews and debriefings were audiotaped.

The "Self-validation" Procedure

The self-validation protocol was based on previous research into the cognitive processes of questionnaire response. Lessler, Bercini, Tourangeau, and Salter (1986) assert: "One consistent finding of our studies, survey research in general, and cognitive science is that recall is improved if:

1. multiple recall cues are given,
2. more time is allowed for the recall of events, and
3. the respondent puts more effort into the recall process."

Additionally, as Martin (1987) has pointed out, retrospective surveys concerned with ambiguous or subjective concepts place a significant demand on the respondent to understand the kind of information the survey is trying to collect. "Last week" and "work" are such concepts. Accurate recall demands that respondents understand how these terms are being used in the survey.

We began by explaining the task to the respondent. Then, the respondent was asked to produce recall cues for the reference week. Recognizing that the best recall cues for a person are events that affected him or her directly (Baddeley et al., 1978), we had the respondent look at a calendar containing the week of interest and provide salient events for each day of this week, such as "worked late," "saw a movie," "went to the mall." The interviewer or respondent wrote these events on the calendar. This time-consuming process required significant effort on the respondent's part, and sensitized him/her to the recall effort that was expected.

After these idiosyncratic recall cues were recorded, the interviewer explained to respondents what was meant by hours worked, presenting examples of activities to be included and excluded. The interviewer then proceeded to ask specific questions about the respondent's work experiences during the reference week. With their idiosyncratic cues in front of them for their week of interest, respondents were asked when they came in to work, when they left, how much time they took off for lunch and other personal activities, about commuting time and other work related travel, and about other work-related activities in which they engaged.

These specific questions began with the most recent day in the week of interest, since it has been shown that most people recall better if they begin with the most recent item in a series and work backwards (Whitten & Leonard, 1981; Williams & Hollan, 1981; both cited in Bradburn, Rips, and Shevell, 1987). Since some people prefer to recall events in the other direction (Loftus & Fathi, 1985), respondents' preferences were accommodated. These questions were repeated for each day of the week. Respondents were encouraged to look at their idiosyncratic calendars to help them answer these specific questions for each day. The interviewer totaled each day's hours worked, then totaled the week's hours. The debriefing interview and worksheet are included in Edwards, Levine, and Allen (1989).

Treating data obtained from the debriefing as "truth" permitted assessment of error magnitude and direction for each of the questionnaire variations being tested and determination of the relative accuracy of self and proxy reporting of hours worked.

Although we believe the number of hours worked elicited by this debriefing is a "truer" measure of hours worked than could be obtained from other sources, others might argue that this measure is not necessarily more valid than the individual's initial report. Accordingly, we tested the validity of the debriefing's measure of hours worked in three ways:

- Individuals were asked if the total number of "debriefed" hours they worked seemed correct, and were asked to help identify the source of all discrepancies between this report and their initial report;
- If another household member was concurrently interviewed, this proxy respondent was shown a detailed, day-by-day breakdown of the hours worked (produced through the debriefing with the self-respondent), and was asked to help identify the source of all discrepancies between this report and his/her initial report; and
- Individuals were asked about the component of hours worked that could be "verified" by their employers, and asked for permission to contact their employers to verify this component. (When this component differed from the number of hours they reported working, they were asked for the reasons for this discrepancy.)

Results

Two different measures of error were calculated for each self- and proxy report -- net error, defined as the questionnaire report of the number of hours worked minus the number of hours worked reported by the respondent in their debriefing, and absolute error, the absolute value of the net error. (Due to improper questionnaire administration, data for five subjects were excluded from the analyses.)

Table 1 summarizes results for absolute error, that is, the absolute magnitude of discrepancies between questionnaire reports and "validating" self-reports. These results confirm the intuitive expectation that self-report of hours worked is more accurate than reporting about another person. The mean absolute error associated with proxy-report (5.68 hours) was significantly greater than self-report error (2.78 hours) ($t=4.01$, $p=.0002$). Also, the variance associated with proxy error was much greater than that associated with self-report error.

Results for net error, which considers both magnitude and direction of error (over- or under-reports), are

summarized in Table 2. Overall, self-respondents over-reported the number of hours they worked by an average of 1.34 hours. This over-reporting "error" is significantly different from 0 ($t=2.46$, $p=.017$). Proxy respondents showed a slight tendency to under-report the number of hours worked by other household members, but this under-reporting was not significantly different from 0 ($t=0.42$, $p=.67$).

We validated the debriefing procedure first by asking respondents if the total number of "debriefed" hours they worked seemed correct, and to help identify the source of all discrepancies between this report and their initial report. This validation step indicated the procedure had face validity and that reasons for discrepancies could be ascertained. When the debriefed number of hours was checked with the other concurrently interviewed household member, the reasonableness of the revised estimate was verified in nearly all cases. In some rare cases, the proxy strongly disagreed and might have been correct; the respondent might have been in error. However, we have assumed that such situations are the exceptions, and that the respondent's assessment of his/her own hours worked, based on greater knowledge, is more likely to be true.

Each employed (but not self-employed) subject was asked for written permission to contact his or her employer. This employer contact served two purposes. It served to "validate" a component of total hours worked, and it also helped to "validate" the self-validation methodology described above. The debriefing provided a detailed, day-by-day breakdown of hours worked, permitting identification of the hours that could be verified from employers' records. If these hours were verified by employer records, it would suggest that other hours that could not be verified through employer contact but were produced by the same techniques and procedures were comparably "valid" measures.

To the extent that the work activities for which hours were reported by the respondent and activities for which the employer had records of hours worked coincided, we considered the record check a complete validation. The employer data would also validate leave hours, paid overtime, and similar deviations from a "normal" work week. Of course, employers would not be likely to have records of extra unpaid hours worked at home, for example. To address this problem, we asked respondents in the debriefing how many hours they thought their employers would report and, when appropriate, why this number was different from the number of hours they actually worked. Thus, the employer validation was intended to "match" definitionally a specific subset of hours reported by the self-respondent.

Of the 71 subjects, 7 were self-employed and 5 were unemployed. Of the remaining 59, 42 (71%) gave their written consent for employer contact. Interviews were completed with 38 (90%) of the employers. Of these, 29 had records of hours worked for the subjects; in the remaining 9 cases, the employers did not keep records.

Of the 29 subjects for whom employer records were available, 17 (59%) predicted exactly what the employers reported. For 3 other subjects (10%), an explanation of the difference was provided by either the employer or subject that indicated exact agreement between the two. (For example, one subject worked half-time over two weeks, three days one week and two days the next. The subject predicted the employer would report the week's actual hours (21), but the employer reported half-time (17.5 hours) for the week.) Of the 9 (31%) remaining subjects, discrepancies ranged from -2.5 to +2 hours. The mean net error (employer report minus subject prediction) was -.12 hours, as opposed to -1.34 hours comparing the self-validation and questionnaire reports; the mean absolute error (absolute value of employer report minus subject prediction) was .47 hours, as opposed to 2.78 hours comparing the self-validation and questionnaire reports.

Conclusion

There was considerable evidence that our debriefing protocol was successful. First, we found close agreement between the debriefed self responses and employer reports for the hours that could be validated in this way. This agreement supports the contention that the debriefing reports of other hours worked were comparably accurate. Second, when debriefing-derived data were used to compare the error of self and proxy questionnaire reports, considerable self-report error was detected, and the intuitive expectation was met that errors by proxies would be larger and more variable than those by self-respondents. Finally, the estimates obtained from the debriefing seemed reasonable to the self-respondents, and more accurate than their initial questionnaire reports.

B. The use of a debriefing protocol for decomposition of reporting error into determinant categories

Introduction

In order to accurately answer a survey question, the respondent must:

- understand what is being asked,
- be sufficiently motivated to undertake the response task,
- know (have in memory) the information requested,
- be able to retrieve appropriate information from memory, and
- formulate and provide a response to the interviewer.

A failure in any of the above processes will introduce error into the measurement process (Biderman et al., 1986; Tourangeau, 1984; Cannell et al., 1981). Response error, therefore, can be seen as a function of five factors -- respondent comprehension, motivation, knowledge (whether the needed information is in memory), retrieval, and response formulation or selection. These factors, which may operate singly or in combination in a survey, provide a framework for evaluating alternative questionnaire designs and for the classification and analysis of survey errors.

Recognizing the influence of these factors, alternative questionnaire versions and administration procedures were developed to address specific aspects of the response process. For example, one alternative provided the respondent with explicit definitions of crucial concepts to enhance comprehension; another provided a calendar to respondents to enhance both memory and motivation.

Besides assessing the effects of these alternatives on net and absolute error, we wished to investigate their effects on specific components of error corresponding to each of the factors listed above. The procedures developed to accomplish this and results of their application comprise the methods and results sections.

Methods

As part of the debriefing protocol, respondents were asked to help reconcile all discrepancies between their original questionnaire self-report answers and the number of hours worked that resulted from the debriefing. Interviewers recorded the respondents' explanations for each discrepancy detected.

Based on these explanations, completed questionnaires, completed debriefing protocols, and reviews of audiotapes (when necessary), senior project staff coded the presence, magnitude and direction of each discrepancy according to categories corresponding to the type of error. (In certain cases, only the presence of a specific type of error could be determined. The magnitude could not be ascertained. These cases are included in calculations of incidence but not of magnitude of error.)

Two kinds of variables were created to describe classified errors: a set of dummy variables indicating the

presence or absence of a type of error in a questionnaire self or proxy report, without regard for the direction of the error; and a set of continuous variables, indicating the magnitude (in hours) and direction (over- or under-report) of each error. If a person made two errors of the same kind in different directions, each error would be included in the appropriate continuous variable and the dummy variable would be set to "1." For example, suppose in his questionnaire report a respondent forgot one hour of sick leave and two hours of overtime, and made no other errors. The **net error** would be -1, or an under-report of one hour. Two hours of memory error under-report would be coded, along with one hour of memory error over-report. The dummy variable for memory error for this person as self-respondent would be set to "yes." Errors of undetermined magnitude were coded as "yes," but no value was added to the relevant continuous variable. In such cases, the entire net error was considered as unclassifiable for calculating magnitude.

At least 10 percent of the questionnaires were independently classified by two staff members. Discrepancies were discussed and resolved. A further review of tape recordings of approximately 25 percent of the questionnaires insured that at least 35 percent of the coding was done at least twice. About 15 percent of all discrepancies were deemed not classifiable. A formal evaluation of inter-coder reliability was undertaken after the fact, yielding a Kappa of .56 with a standard error of .22, indicating fair to good agreement.

Results

Table 3 shows the incidence and mean magnitude of all errors detected in comparing questionnaire reports by self and proxy respondents with the self-validation reports. The incidence of errors (number of errors detected / number of respondents) is given for over-reports, under-reports, and in total. Overall, an average of .93 errors were detected per self-respondent and 1.25 errors per proxy respondent. (Note that a single respondent could have been, and often was, credited with more than one error.) The incidence of over-reporting was .51 for both self- and proxy respondents, while .42 errors of under-reporting were detected per self-respondent and .74 per proxy respondent.

"Mean magnitude" in Table 3 refers to the average magnitude of errors among respondents who made them. Thus, the average magnitude of an over-reporting error by a self-respondent was 4.82 hours, while for a proxy respondent it was 5.90 hours. For both self- and proxy respondents, the mean magnitude of over-reporting errors was considerably higher than for under-reporting -- 48 percent higher for self-respondents, and 32 percent for proxy respondents. The combined effects of incidence and magnitude are, of course, consistent with the net error results described earlier: self-responses appear biased toward over-reporting, while proxy responses show no evidence of bias but are subject to more error.

Table 4 expands the calculation of incidence and mean magnitude of errors across the five error types described earlier. The tendency for over-reporting by self-respondents persists throughout each error type, while there is no such discernible directional pattern for proxy respondents. Both in incidence and magnitude, motivation and comprehension errors dominate for self-respondents, while knowledge errors are by far the most common and result in the greatest contribution to absolute and net error for proxy respondents. The incidence of memory error and unclassifiable error is nearly identical for self- and proxy respondents, with proxies tending to make larger errors in these categories.

The difference between self- and proxy reports in comprehension and motivation error is striking -- one would expect no difference in level of understanding or motivation depending upon whether a respondent is answering for him- or herself or for another person. The reason for the difference is partly in the definition of the error terms. Self-respondents were assumed to have perfect knowledge -- that is, given proper definitions, a perfect recall strategy, and perfect motivation, self-respondents would report the truth in the debriefing. If a proxy respondent made an estimate of (rather than attempting to calculate exactly) the number of hours a subject worked, and said in the debriefing that he or she simply did not know how many hours the subject worked, any error would have been attributed to lack of knowledge. If a self-respondent made such an estimate, and in the debriefing came up with a more exact figure, the error would have been classified as motivation (if no apparent attempt to remember the week's hours was made) or memory (if there was some indication of such an attempt). Thus, some proxy motivation error may have gone undetected when the proxy simply pleaded ignorance.

Similarly, greater knowledge by self-respondents may account for the difference in comprehension error between self- and proxy reports. To make a comprehension error, a respondent would have to have made an attempt to recall hours worked exactly, or to have applied a definition different from the CPS meaning to a usual week. Such errors were frequently small or made up of small components (such as a half hour of unpaid set-up time repeated over all work days).

Conclusion

Using the debriefing protocol described in the first part of the paper, we were able to determine the magnitude and direction of discrepancies between questionnaire reports and "validation" totals. The classification scheme described in this section allows further analysis of these discrepancies according to the reason for error. Generally, investigation of the causes of response error can prove very useful for survey designers in the identification of specific types of problems associated with questionnaire wordings as well as suggesting specific approaches to be taken in the amelioration of these problems. To intervene at the level of comprehension, for example, one might devise methods of giving the respondent more complete definitions of terms like "hours worked." If motivation is perceived as a serious problem, the design can intervene directly at this level (by including motivational statements about the importance of accurate answers, for example) or can focus instead on offering respondents assistance with an estimation strategy. Offering an effective estimation strategy may also be appropriate if one is attempting to mitigate the effects of knowledge error. Alternatively, one could limit the selection of proxy respondents by some criteria or use a record check or multiple respondent approach. Response selection error has many and varied potential causes; some motivation interventions might also ameliorate response selection error, as could the use of proxy responses or validation from records. Of course, an intervention aimed at reducing one kind of error could aggravate another kind. For example, using proxy respondents to limit response selection error could increase knowledge error.

In the research described here, it appears that lack of knowledge is the most important factor contributing to proxy error, while lack of motivation is the primary reason for errors by self-respondents, with differences in definitional interpretation (comprehension error) also contributing significantly to errors by self-respondents. We are fairly confident that these kinds of conclusions may be drawn from this classification system and analysis. We are less confident of other kinds of uses of the classification information. For example, we would not be justified in drawing conclusions about the relative motivation of self and proxy respondents by comparing the incidence of motivation error between the two.

In our report to BLS, we have suggested that changes to

the CPS questionnaire and procedures (using a calendar and recall cues, limiting the use of proxy respondents) be considered for further testing as a result of this research. However, these kinds of changes have significant implications for the level of resources required to complete the CPS, and also potentially for the tight production schedule to which the CPS adheres.

REFERENCES

Baddeley, A., Lewis, V., Nimmo-Smith, I. (1978), "When did you last...?" in Practical Aspects of Memory, eds. M.M. Gruneberg, P.E. Morris, and R.N. Sykes, London: Academic Press.

Biderman, A., Cantor, D., Lynch, J., and Martin, E. (1986), Final Report of Research and Development for the Redesign of the National Crime Survey, Bureau of Social Science Research, Washington, DC.

Bradburn, N., Rips, L., and Shevell, S., "Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys," Science, 1987, *236*, 157-161.

Cannell, C., Miller, P., and Oksenberg, L. (1981), "Research on Interviewing Techniques," in Sociological Methodology, ed. S. Leinhardt, San Francisco: Jossey-Bass, pp. 389-437.

Edwards, W.S., Levine, R., and Allen, B. (1989), "Research on Hours of Work Questions in the Current Population Survey." Final Report, Bureau of Labor Statistics Contract No. J-9-J-8-0083.

Lessler, J., Bercini, D., Tourangeau, R., and Salter, W. (1986), "Cognitive Laboratory Studies of the 1986 Supplement to the National Health Interview Survey Final Results," in Proceedings of the Section on Survey Research Methods, Washington, DC: American Statistical Association, pp. 478-480.

Loftus, E., and Fathi, D. (1985), "Retrieving Multiple Autobiographical Memories," Social Cognition, *3*, 280-295.

Marquis, K.H. (1984), "Record Checks for Sample Surveys," in Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines, T. Jabine, et al., Washington, DC: National Academy Press.

Martin, E. (1987), "Some Conceptual Problems in the Current Population Survey," in Proceedings of the Section on Survey Research Methods, Washington, DC: American Statistical Association, pp. 420-424.

Tourangeau, R. (1984), "Cognitive Sciences and Survey Methods," in Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines, T. Jabine, et al., Washington, DC: National Academy Press.

Whitten, W.B., II, and Leonard, J.M. (1981), "Directed Search through Autobiographical Memory," in Memory and Cognition, *9*, 566-579.

Williams, M.D., and Hollan, J.D. (1981), in Cognitive Science, *5*, 87.

Table 1

Mean Absolute Error by Questionnaire Variation

		Self-Reports			Proxy Reports		
		N	Mean	S.D.	N	Mean	S.D.
Variation	K	15	3.82	4.67	11	6.06	5.61
	L	16	2.57	3.26	10	2.54	2.13
	M	18	1.85	3.15	14	6.01	7.78
	N	17	3.04	3.61	14	7.29	6.28
All Variations		66[1]	2.78	3.67	49[2]	5.68	6.11

[1] 5 if 71 subjects were not employed

[2] 5 subjects of 56 with proxy responses were not employed; 2 interviews were discarded because of improper administration

Table 2

Mean Net Error by Questionnaire Variation

		Self-Reports			Proxy Reports		
		N	Mean	S.D.	N	Mean	S.D.
Variation	K	15	3.13	5.19	11	0.48	8.46
	L	16	1.59	3.87	10	0.44	3.39
	M	18	0.38	3.66	14	1.97	9.76
	N	17	0.54	4.75	14	-4.45	8.67
All Variations		66	1.33	4.41	49	-0.51	8.37

Table 3

Incidence, Magnitude, and Direction of Discrepancies (Errors) between Questionnaire Reports of Hours Worked and "Validated" Self-Reports, for Self and Proxy Respondents

			Over-Reports		Under-Reports	
	N	Total Incidence[1]	Incidence[2]	Mean Magnitude[3]	Incidence	Mean Magnitude
Self-Report	66	.93	.51	4.82	.42	3.26
Proxy Report	49	1.25	.51	5.90	.74	4.46

[1]Total incidence = (number of over-report errors detected + number of under-report errors detected)/N

[2]Incidence = (number of over-report errors detected)/N.

[3]Mean magnitude (in hours) = (Σ over-report errors [in hours])/number of respondents making over-report errors.

Table 4

Incidence, Magnitude, and Direction of Discrepancies Between Questionnaire Reports of Hours Worked and "Validated" Self-Reports, by Type of Error

	Self-Report (n = 66)					Proxy Report (n = 49)				
		Over-Report		Under-Report			Over-Report		Under-Report	
	Total Incidence[1]	Incidence[2]	Mean Magnitude[3]	Incidence	Mean Magnitude	Total Incidence	Incidence	Mean Magnitude	Incidence	Mean Magnitude
Comprehension	.26	.12	6.67	.14	1.64	.16	.04	12.50	.12	4.67
Memory	.17	.11	3.00	.06	3.17	.22	.06	10.00	.16	4.19
Motivation	.37	.17	6.58	.20	4.65	.08	.02	4.00	.06	3.83
Knowledge	-	-	-	-	-	.57	.33	4.00	.24	6.54
Response Selection	.02	.02	4.00	0	0	0	-	-	-	-
Unknown	.11	.09	1.44	.02	1.00	.22	.06	8.50	.16	3.12
Total	.93	.51	4.82	.42	3.26	1.25	.51	5.90	.74	4.46

[1]Total incidence = (number of over-report errors detected + number of under-report errors detected)/N.

[2]Incidence = (number of over-report errors detected)/N.

[3]Mean magnitude (in hours) = (Σ over-report errors [in hours])/number of respondents making over-report errors.