ANSWERING QUESTIONS, QUESTIONING ANSWERS: EVALUATING DATA QUALITY IN AN ESTABLISHMENT SURVEY

Karen L. Goldenberg, Bureau of Labor Statistics
2 Massachusetts Avenue N.E., Room 4985, Washington, DC 20212

KEY WORDS: Data quality; measurement error; establishment survey; response analysis survey.

1. Introduction: Survey Measurement Error

Every survey, by definition, asks respondents to answer questions. As the role of surveys has grown in importance in everyday life, we as researchers have come to ask questions ourselves. Our objective in conducting a survey is to find some "truth," and we want to know whether the data our respondents give us are coming close to that truth. So we question respondents' answers. We ask about their responses and the circumstances under which they give those answers. That is, we look at survey measurement error.

This paper examines selected aspects of measurement error for an establishment survey. Although the measurement error literature is extensive, the Federal Committee on Statistical Methodology noted in 1988 that "very little in the way of theoretical or evaluative work on survey quality has been published for establishment surveys" (p.1). This situation is changing,[1] but the basis of most measurement error research to date is surveys of individuals or households, with emphasis on reported attitudes and behaviors. Groves (1989) identifies four sources of measurement error: the respondent, the questionnaire, the interviewer, and the mode of data collection, all of which may be present in establishment as well as household surveys. However, establishment surveys differ from household and individual surveys in the type of data collected and the respondent's role in providing those data, and so can have additional sources of error.

An establishment survey is a "census or sample survey whose sources of information are public or private businesses, agencies, or other nonhousehold organizations, or individuals acting as representatives of them" (Edwards and Cantor, 1991: 212). Establishment surveys are often job-related, and interrupt or are imposed upon the respondent's working day. Questionnaires for establishment surveys may contain professional terminology or jargon (Phipps et at., 1993), and respondents answer questions that relate to their organizational roles. Although locating the correct respondent can be critically important to the quality of responses obtained (Dutka and Frankel, 1991), analysis of survey data is usually based on characteristics of the establishment rather than those of the respondent (Moore and Baxter, 1993). Finally, establishment surveys typically seek "hard data" or information from organization records (Federal Committee on Statistical Methodology, 1988). This is especially true for surveys conducted by or for government agencies.

Dutka and Frankel (1991) distinguish between *analytical* and *enumerative* establishment surveys. Analytical establishment surveys are similar to household surveys, in that they ask about a respondent's attitudes and behaviors, albeit in an organizational context. Groves' four sources of error apply. For example, the **respondent** must read or hear and understand each question and formulate a response to it, drawing on his or her personal knowledge, experience, and opinions. The respondent may make errors in recall or estimation, misplace events in time, or misunderstand a question. The **questionnaire** may introduce errors through poorly worded questions or inadequate layout and instructions. The **interviewer** may make errors in conducting the interview. **Mode** effects may result if some respondents complete self-administered questionnaires and others answer identical questions in telephone or personal interviews, because of the different sets of communications methodologies involved (Groves, 1989). As long as a knowledgeable person is responding to the survey, measurement error properties for analytical establishment surveys should be similar to those for household surveys.

Enumerative establishment surveys, on the other hand, measure characteristics of the establishment as a whole, based on information from establishment records. Government surveys and censuses often fall into this category. Enumerative surveys are frequently conducted by mail, using self-administered data collection instruments. Because the focus of the data is the establishment, **establishment records** comprise an additional source of measurement error for these surveys. Records as an error source are imposed upon the sources noted above.

Establishment records contribute to measurement error through characteristics of the organization and of the respondent vis-a-vis the organization. Attributes of the establishment, such as size and industry, shape the magnitude and complexity of the record-keeping system, while those of the person(s) who respond for the establishment, such as position in the organization, knowledge of the subject, and familiarity with the appropriate records, affect the extent to which the correct data are reported (Goldenberg et al., 1993). Although the respondent may understand the question perfectly well, the establishment's records may not contain the needed information, the data may not be aggregated in a way that meets survey requirements, or

[1] The International Conference on Establishment Surveys, held in Buffalo, New York in June 1993, was an attempt to address this imbalance by bringing together researchers involved with survey methods for businesses, farms, and institutions.

the respondent may not have the knowledge necessary to prepare the answer.

This paper looks at measurement error in the Hours at Work Survey (HWS), an ongoing establishment survey conducted by the Bureau of Labor Statistics (BLS). The approach used to study measurement error is a Response Analysis Survey (RAS), essentially a respondent debriefing in which we recontact survey respondents after they submit completed mail questionnaires. A RAS complements cognitive pretesting techniques by using a structured questionnaire and generating quantitative data. Since it is administered to a subsample of respondents from the original survey, results can be generalized (Goldenberg et al., 1993).

The paper describes the Hours at Work Survey and the 1992 Response Analysis Survey for the HWS. It then turns to the results of the RAS, and shows how the process of questioning answers offers insight into survey data quality.

2. The Hours at Work Survey

The Hours at Work Survey (HWS) is a national survey of business establishments that has been conducted by the Bureau of Labor Statistics (BLS) each year since 1982. Using the Dutka and Frankel (1991) terminology, it is an enumerative survey. The objective of the survey is to obtain inputs to measures of productivity. It does so by collecting from each participating establishment the number of hours it paid production or nonsupervisory employees the previous calendar year, and the number of hours those employees actually worked (i.e., hours paid minus paid leave). BLS uses the results of the survey to compute a ratio of hours worked to hours paid (HW/HP), which it applies to data from the Current Employment Statistics program to generate productivity statistics.[2]

The HWS sample consists of approximately 6,000 private nonagricultural establishments in the 50 States and the District of Columbia that report employment and earnings to State unemployment insurance programs. Sample establishments receive the HWS schedule, a one-page form-based questionnaire. The questionnaire requests the total hours paid and hours at work for the previous year, and for each quarter of that year, and contains questions about the number of employees and types of paid leave offered.

There are two versions of the questionnaire, one for production workers (in manufacturing, mining, and construction establishments) and one for nonsupervisory workers (establishments in all other industries). The primary difference in the form is the reference group of workers for which it collects data.

The survey mailing takes place early in the calendar year. Survey procedures call for an initial mailing and two mail follow-ups to nonrespondents. A few weeks after the third mailing, HWS staff interviewers begin to contact nonrespondents by telephone in an attempt to obtain the data. If the respondent says that the data necessary to answer the questions are not available, the interviewer works with the respondent to complete a worksheet that results in estimates of hours paid and hours at work. The worksheet obtains somewhat different information from the HWS form: the number of production or nonsupervisory workers, the *average work week* for a production or nonsupervisory worker, including overtime, and the average amount of paid leave received by production or nonsupervisory workers. A computer algorithm uses average work week and number of production or nonsupervisory workers to estimate hours paid, and subtracts paid leave to determine hours at work.

Combining the mail and telephone data, response rates for HWS are generally at or above 70 percent, and survey procedures ensure that a 70 percent response rate is achieved for each industry and size class combination in the sample. The overall rate for 1992 was 74.7, with two-thirds (66.4 percent) obtained by mail or fax and the remaining third collected by telephone. The 1992 survey yielded a total of exactly 4,000 usable responses (Goldenberg, 1993).

Some BLS analysts have expressed concern about the practice of combining the mail responses with the estimates provided by telephone respondents (Barkume, 1990), fearing a biasing effect on the HW/HP ratios. We compared HW/HP ratios for mail and telephone respondents and found that differences between them were random (without biases in either direction) and not statistically significant. Nevertheless, this concern was one factor motivating the 1992 RAS.

3. The Hours at Work Response Analysis Survey

3.1 Background

The 1992 RAS was the first effort since HWS' inception in 1982 to evaluate the quality of the data. Over the years a number of changes have taken place. While intended as a mail survey, the staff began to follow up nonrespondents by telephone, raising some of the concerns noted above. Changes in the external environment have resulted in lower mail survey response rates, and response for the industries with nonsupervisory workers has been consistently lower than that for other industries. At the same time, the movement of personal computers into business operations led BLS to expect the hours data requested in the HWS would become more readily available.

3.2 HWS RAS Objectives and Procedures

The 1992 HWS RAS objectives included: 1) An evaluation of the quality of the HWS data. "Quality" here refers to the extent to which responses come from records and conform to BLS definitions. 2) An examination of the availability of records and data summaries from which respondents obtain hours data, both as an indicator of data quality and as a measure of response burden. 3) An assessment of differences between HWS participants who respond by mail and

---

[2] The Current Employment Statistics program is a monthly BLS establishment survey that monitors the movement of jobs.

those who only respond after being contacted by telephone, and if their reporting practices differ, how.

We began by drawing separate subsamples of HWS establishments who replied by mail or fax and establishments that responded only after being contacted by telephone. The final sample consisted of 290 establishments from the mail response group and 294 from the telephone group, and yielded a total of 458 interviews.

The RAS consisted of a telephone interview lasting about 10 minutes with the individual who actually completed the form. The staff was more successful in interviewing HWS mail respondents (N=273) than it was with members of the telephone group (N=185). The response rate for the mail sample was 94 percent, and for the telephone respondents was 63 percent, for an overall response rate of 74 percent.

Following data collection, we compared RAS respondents and HWS respondents and found them to be generally similar. RAS respondents are from all industries in the HWS sample, from all size classes, and from every state in the United States. The biggest difference between the two sets of respondents is that two-thirds of the HWS respondents were from production worker establishments, while the RAS respondents were more equally divided between production and nonsupervisory worker firms.

Interview topics included sources of information for reporting hours data, and availability of that information in summary reports; employee work schedules (full time, part time, temporary or seasonal); types of leave and how that leave was reported on HWS; groups of employees included in the hours data; overtime reporting. These questions allowed us to put the HWS data into context. For example, questions on the availability of data summaries helped to clarify the magnitude of the respondent's task in preparing the HWS, and showed where errors might be introduced.

4. HWS Data Quality

4.1 Measurement Error and Data Quality

The HWS RAS concentrates on data quality issues associated with the questionnaire, the establishment's record-keeping system, and the respondent's adherence to BLS definitions in compiling answers to the questions. We assess error by looking at the percentage of HWS respondents who perform the task as intended.

An important source of measurement error in many surveys is that of *mode*, and HWS is no exception. However, we cannot speak of a true mode effect here, because the nonresponse follow-up procedures collect data using different questions, and RAS evidence suggests differences in the underlying populations. Also, the telephone interviews have not been formally structured, so errors associated with them are almost impossible to measure.[3]

4.2 Data Sources: Data Availability

One measure of data quality concerns whether an establishment has records that respondents can consult. Are the data present in the employer's files, and if so, in what form? The more closely the data conform to survey definitions, the lower the response burden, the higher the response rate, and the higher the data quality. If an establishment has to aggregate a large amount of information, it may not respond to the survey; it if does respond, the potential for calculation and other errors increases with the effort.

The RAS addressed the data availability issue with parallel sets of questions about data for hours paid and hours at work or paid leave. In order to complete the HWS form, a firm must have basic data on hours paid and hours at work or hours of paid leave for each individual employee. Therefore, we asked if the establishment kept these individual records, and if so, whether the firm summarized the records either quarterly or annually. We also asked whether the establishment produced a summary or report, for its own purposes, of the total hours paid, or of the total hours at work (or paid leave). If the respondent said the establishment had a summary report, we asked if it was produced quarterly or annually, and whether the report showed hours for production or nonsupervisory workers only, or if it included other employees.[4]

Overall, more than four-fifths of respondents said that they kept records of individual employees' hours paid and hours at work or paid leave. About half that number prepared summaries of individual employee hours data by quarter or for the year. Having data summaries by individual workers makes it possible to tabulate information limited to production or nonsupervisory workers, but a substantial amount of effort could be involved.

The data show noticeable differences between mail and telephone respondents. Mail respondents are almost twice as likely as telephone respondents to have summary data for individual employees, both by quarter and for the year. These differences are statistically significant at the $p < .001$ level.[5] There are no differences based on industry (production/nonsupervisory or manufacturing/nonmanufacturing) in terms of preparing individual hours summaries.

---

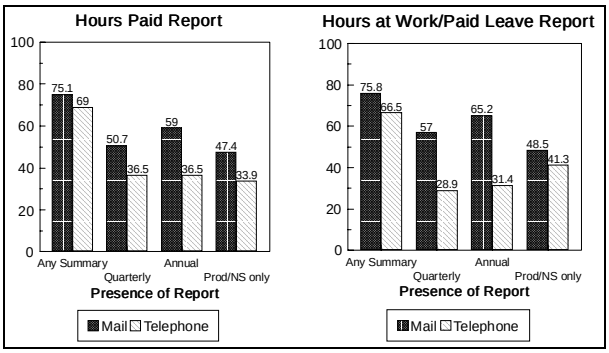
[3] We should note here that the RAS is subject to the same types of errors as any other survey. Since we were speaking to previous respondents about their behavior, the RAS was an analytical rather than an enumerative survey. We have no accuracy measures for RAS responses.

[4] We did not investigate adherence to definitions in these summary reports. A 1982 HWS RAS showed that some companies included holidays in reports of regular work hours (Scott, 1983), and subsequent investigation has identified a few special types of paid leave, such as jury duty, that employers may count as regular work hours. These questions should be raised in future research.

[5] We use the chi-square test of homogeneity for determining the differences in proportions of interest. Where differences are statistically significant we show the probability p of incorrectly rejecting the hypothesis that the proportions are equal. When $p > .05$ we report the differences as not significant.

Some establishments aggregate data across employees to prepare summary reports on hours paid and hours at work. Figure 1 shows that there are differences between mail and telephone respondents, with fewer telephone respondents having summary reports. While this analysis does not attempt to establish causation, we can speculate that the telephone respondents' lack of data may be one reason for not responding to the mail questionnaire.

Figure 1. Availability of Summary Reports



*Percent preparing each type of summary report

Quarterly and Annual Summaries. We asked respondents whose firms had hours paid or hours at work summaries if they produced those summaries quarterly, and if they produced them annually. The results are very similar for the two sets of information. Overall, 45-46 percent of the establishments said that they generate quarterly reports, and 50-53 percent have annual reports. While industry and size have little effect, there are large and statistically significant differences between mail and telephone respondents. Fewer telephone than mail respondents prepare data summaries for either hours paid or hours at work.

Again, we can speculate on the relationship between having an annual summary report on hours paid and the likelihood that an establishment will complete and mail back the HWS form. For hours paid, 59 percent of mail respondents produced annual summary reports, compared to 36 percent of telephone respondents ($p < .001$). The difference is smaller but still statistically significant for respondents with quarterly summary reports. Mail-telephone differences are even more pronounced when the subject is annual summaries of hours at work or paid leave ($p < .001$).

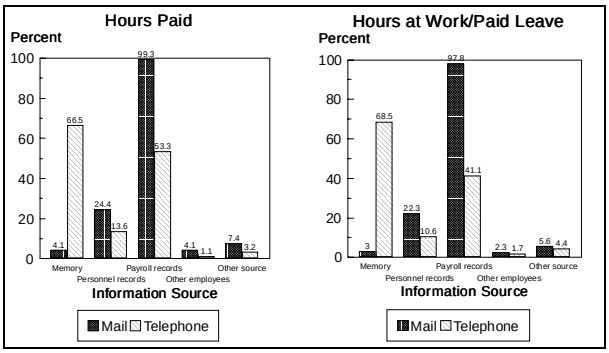
Production or Nonsupervisory Workers Only. Fewer than half of the establishments produce summary reports limited to production or nonsupervisory workers; in fact, only 42 percent have them for hours paid and 46 percent for hours at work or paid leave. Manufacturing industries hiring production workers are more likely to have a report on hours paid limited to workers of interest to HWS (49 percent) than either nonmanufacturing industries with production workers (38 percent) or nonmanufacturing industries with nonsupervisory workers (34 percent). This difference is statistically significant ($p < .05$). A slightly larger percentage of establishments has an hours at work or paid leave report for the population of interest to HWS. Telephone respondents are less likely than mail respondents to have data summaries for production or nonsupervisory workers.

4.3 Data Sources: Use of Records

Another measure of data quality is whether or not the respondent used records to answer the questions, or answered from memory, by asking a coworker, or seeking out another data source. The RAS included questions about the sources of data the respondents used, first for hours paid and then for hours at work, and asked specifically about memory, personnel records, payroll records, other employees, or any other source of information. In this context, memory includes knowledge of the subject matter. Respondents could answer "yes" to any or all of the five information sources.

Results showed that the majority of respondents obtained data from establishment records. At least three-fourths used payroll records, and another 17-20 percent used personnel information. It is clear from Figure 2 that there are important differences between mail and telephone respondents in their sources of information, consistent with the differences in availability of summary data. Mail respondents almost universally consulted payroll records, but a much smaller percentage of telephone respondents did so: just over half for hours paid, and about two-fifths for hours at work/paid leave information. Equally important, no more than 4 percent of mail respondents relied on memory to prepare HWS data. These differences are statistically significant.

Figure 2. Sources of Information Used



*Percent using each data source

4.4 Adherence to BLS Definitions

Another aspect of data quality is whether respondents prepare data according to BLS definitions. The hours paid and hours at work definitions are:

Hours paid. The number of hours for which all production or nonsupervisory workers *received pay* for the previous calendar year. Hours paid includes regular and overtime hours at work, plus all holidays, vacation, paid sick days, jury duty, military leave, and other paid or personal leave.

Hours at work. The number of hours an employee spends on the employer's premises, on duty, or at a prescribed workplace. Hours at work includes normal working hours, travel time between job sites during the workday, stand-by time, and rest periods. Hours at work does not include any paid leave.

We looked at several areas where respondents could deviate from BLS definitions in preparing answers to the HWS, including:

Reporting paid leave. Does the establishment offer paid leave to full time employees? If so, was paid leave included in hours paid and excluded from hours at work? Was the amount of paid leave the actual amount used by employees, or the amount employees were entitled to use?

Reporting overtime. Were overtime hours included in the hours paid and hours at work data? If so, were they counted as the actual number of hours, or equivalent to some premium pay factor?

Including appropriate employees. Did the establishment include only production or nonsupervisory workers in the hours data, or did it include managers or others that the definition specifically excludes?

4.4.1 Reporting Paid Leave

Paid leave is an important component of the hours paid and hours at work variables. By definition, if an employer offers its employees paid leave, the total hours paid should include that leave, and hours at work should exclude it. But employers offer different types and amounts of leave to different employee groups. Full time workers are more likely to receive paid leave, and more types of paid leave, than permanent part-time workers or employees brought on for short-term or seasonal work. The effect of handling paid leave incorrectly on HWS data depends on the mix of workers and the types of paid leave offered to those workers.

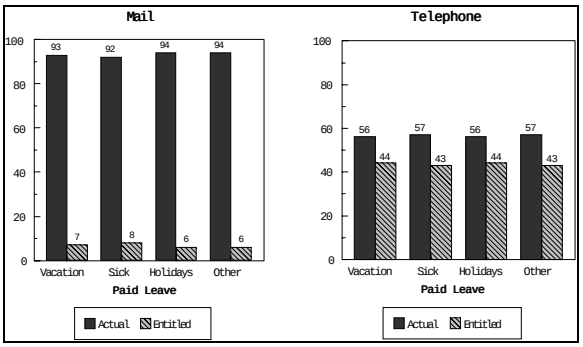
Availability of Paid Leave. Over 90 percent of full time workers receive paid vacation and paid holidays, while 59 percent receive paid sick leave and 48 percent receive other paid time off. About half as many part-time workers receive each type of leave, while temporary and seasonal workers receive almost none. There was very little difference between mail and telephone respondents in terms of the types of paid leave establishments offered their employees.

Treatment of paid leave on HWS form. There are two dimensions to the treatment of paid leave: whether it is excluded from reported hours at work, and whether the amount excluded is the amount actually used or the amount to which an employee is entitled. The correct procedure is to work with the actual amount of leave taken. From 85 to 88 percent of respondents treated all types of paid leave correctly for full time employees. Differences between mail and telephone respondents were negligible. We found mail/telephone differences in the treatment of paid leave for part time employees, but the effect of this error is minimal because few establishments offer paid leave to part time employees. However, mail and telephone respondents operated in different ways in reporting actual or entitled amounts of leave for full time employees. Figure 3 compares these two groups.

Clearly, the vast majority of mail respondents correctly excluded the actual number of hours. However,

telephone respondents are asked for average amounts of paid leave for each of several different leave categories. Almost half of the telephone respondents based this average amount on the amount of leave to which an employee is *entitled* rather than the amount employees actually took in the previous year. Unfortunately, we have no data on the amounts of entitled and actual leave with which to estimate the magnitude of error.

Figure 3. Treatment of Leave in Hours at Work Data, Percentage Using Actual versus Entitled Hours, Full-Time Employees*



* All mail/telephone differences p < .001

4.3.2 Reported Overtime

Another area where we can look at respondent compliance with BLS definitions is that of overtime. Overtime is a component of hours paid and hours at work. If an establishment's production or nonsupervisory employees work overtime, the firm should include those hours in both the hours at work and hours paid figures. Furthermore, if an employee receives premium pay for overtime hours, the number of hours reported should be the actual number of hours worked, rather than the straight-time equivalent of premium-paid hours. The HWS form instructs respondents to count one overtime hour as one hour, even if it is paid at a premium rate.

Paying Overtime. Overall, 88 percent of responding establishments paid overtime to their production or nonsupervisory employees. There was virtually no difference between mail and telephone respondents on this question. Manufacturing firms were more likely to pay overtime to their production workers—96 percent responded to this question in the affirmative—than nonmanufacturing firms with production workers (84 percent) or nonmanufacturing establishments with nonsupervisory workers (80 percent), a difference that is statistically significant (p = < .001).

Reporting Overtime. There were some important mail/telephone differences in reporting overtime. While 95-96 percent of the mail respondents said they included overtime hours in the hours paid and hours at work data, only three-fourths of the telephone respondents reported doing so. Nonmanufacturing firms employing nonsupervisory workers were the most likely telephone respondents to omit overtime from the average work week (33 percent failed to include it in the average work week), although 19 percent of manufacturing firms also omitted it.

About four-fifths of the mail respondents counted one overtime hour as one hour worked, with the rest counting overtime hours as more than one hour (probably the straight-time equivalent of the number of hours worked). Again, there were industry differences, with nonmanufacturing firms employing nonsupervisory workers less likely than employers of production workers to report overtime correctly.

Estimates or Records. For the most part, respondents used records to determine information about overtime: 77 percent indicated that they took overtime hours from records, while the remainder estimated these hours. Virtually all of the mail respondents' overtime data came from records (97 percent), but only 35 percent of the telephone respondents consulted records for these data (p < .001).

5. Discussion

Questioning answers, or asking respondents questions about the context surrounding their answers, is a productive means of evaluating data quality. The Response Analysis Survey for the Hours at Work Survey yielded insights into the sources of information respondents used to prepare their responses. Among the more noteworthy findings, the majority of respondents used records to compile data and answer HWS questions, but there were substantial differences between mail respondents and those contacted during the telephone followup process. Also, there were big differences in the availability of records for mail and telephone respondents. Telephone respondents were much less likely that other establishments to have the types of records that would facilitate completing HWS, a finding which suggests that the presence of records contributes to an establishment's willingness to respond to the mail survey in the first place.

There are industry differences in the availability of records, with manufacturing industries more likely to have hours data for production workers than nonmanufacturing industries are to have data for their nonsupervisory workers. As a result, employers of production workers were better able to report as requested. This is consistent with the pattern of higher mail response rates obtained from manufacturing establishments.

While the RAS looked conceptually at respondents' reporting of information, we made no attempt to assess the validity of the individual numbers reported on the survey form. We cannot identify computational errors, errors made while transcribing information from a summary report, or errors in the establishment's records. Such errors also contribute to measurement error, but they were beyond the scope of this effort.

Where do we go from here? BLS is redesigning HWS data collection form. We are investigating question wording to ensure that respondents understand the intent of each question, and we may ask respondents to tell us if we ask for something that they cannot

provide. We will revise the questionnaire layout to call attention to instructions and to make the form easier to complete.

As part of this effort, the HWS staff is also modifying the survey procedures. This year we experimented with several approaches to increasing mail response, and we will incorporate those that were effective. In addition, we plan to modify telephone procedures, and will attempt to collect the detailed data during nonresponse followup that we obtain by mail. The goal of these changes is to receive more reports based on records. That alone should increase the quality of data. In conjunction with changes to the questionnaire, the new procedures should also increase adherence to definitions. A RAS will follow the introduction of changes to help assess success in meeting our goals.

REFERENCES

Barkume, A. 1990. *BLS Data on Hours at Work: A Discussion Paper.* Internal BLS document.

Dutka, S. and L. R. Frankel. 1991. In Paul Biemer et al, eds., *Measurement Errors in Surveys.* New York: Wiley, pp. 113-123.

Edwards, W. S. and D. Cantor. 1991. In Paul Biemer. et al, eds., *Measurement Errors in Surveys*, pp. 211-233.

Federal Committee on Statistical Methodology. 1988. *Quality in Establishment Surveys.* Statistical Policy Working Paper 15. Statistical Policy Office, Office of Management and Budget.

Goldenberg, K. L. 1993. "Results of the Response Analysis Survey (RAS) for the Hours at Work Survey (HWS)." Internal BLS document.

Goldenberg, K. L., S. J. Butani, and P. A. Phipps. 1993. "Response Analysis Surveys for Assessing Response Errors in Establishment Surveys." *Proceedings of the International Conference on Establishment Surveys.* American Statistical Association, pp. 290-299.

Groves, R. M. 1989. *Survey Errors and Survey Costs*. NY: Wiley.

Moore, D. and R. Baxter. 1993. "Increasing Mail Questionnaire Completion for Business Populations." *Proceedings of the International Conference on Establishment Surveys.* American Statistical Association, pp. 496-502.

Phipps, P., S. Butani, and Y. Chun. 1993. "Designing Establishment Survey Questionnaires." *BLS Statistical Note 35.*

Scott, S. 1983. "Availability of Employee Hours Data." *Proceedings of the Section on Survey Research Methods.* American Statistical Association, pp. 461-465.