

CONVERSION FROM MAIL TO CATI IN THE CURRENT EMPLOYMENT STATISTICS SURVEY

George Werking, Richard Clayton, Richard Rosen, and Debbie Winter
U.S. Bureau of Labor Statistics, 441 G St., N.W., Washington, D.C. 20212

1. INTRODUCTION

Since 1985, the Bureau of Labor Statistics (BLS) has been testing the use of Computer Assisted Telephone Interviewing (CATI) in its monthly establishment employment survey. This research is part of a multi-year project to improve all aspects of the survey. In 1987-1988, BLS conducted CATI production tests in 11 states, following up on an earlier, two state operational feasibility test. This paper contrasts the results of the production tests to the those of the feasibility test and to those currently obtained under mail shuttle collection. Particular attention is paid to the impact of CATI on late respondents and of monthly CATI collection to mail shuttle collection over a two year span. Performance measures include timeliness of response, average number of calls, length of the calls, and attrition. The results show CATI is very effective in improving response rates and other performance indicators.

2. CES SURVEY BACKGROUND

The Current Employment Statistics (CES) Survey is a Federal/State voluntary survey collecting employment, payroll, and hours data on a monthly basis from more than 300,000 nonfarm establishments. The primary mode of data collection since the survey's inception more than 50 years ago has been mail. Each month, participating firms are sent an industry-specific collection form whereby data are entered for the most recent reference period and the form is mailed back to the appropriate State Employment Security Agency. The state agency key enters the data, performs basic edit and screening activities, and transmits the data to BLS to develop national estimates of employment, hours, and earnings. The states use the data to produce state and metropolitan area estimates. The form, used as a shuttle, is then returned to the respondent to enter next month's information.

The collection form requests six data items for the pay period that includes the 12th of the month: total employment; women worker employment; production/nonsupervisory worker employment; production/nonsupervisory worker payroll; production/nonsupervisory worker hours; and either production

worker overtime hours (manufacturing only) or commissions for nonsupervisory workers in trade industries. In the goods-producing industries, data are collected using the production worker designation. In the service-producing industries, the nonsupervisory worker designation is used. In addition, there are special forms used in the government sector, where the data items and definitions differ slightly from those used in the private sector.

The CES Survey has operated in a mail shuttle environment since its inception and strengths and weaknesses of data collection by mail have affected the entire survey process. The primary strength of mail collection is its relatively low cost. Weaknesses of mail shuttle collection are that reporting is at the respondent's discretion and there is lack of contact with the respondent. Furthermore, once the form has been completed and mailed, there is a 1-3 day delay during mail delivery as well as the subsequent time needed by the state agency to key enter and edit the data once received.

As a result of delays in the mail collection process, estimates from the CES Survey are generated three times for each reference month. These estimation points are referred to as "closings". The earliest or first closing estimates are released about 3 weeks after the end of the reference period, based upon 2 weeks of collection. First closing estimates are regarded as the most important, since they are one of the earliest available measures of the nation's economic health, and receive the greatest attention from policy-makers and the public. Under mail shuttle collection, approximately 50% of the sample units report data in time for first closing estimates. As additional responses are obtained, a second set of estimates are produced from an additional three weeks of collection, based on about 75% of the sample. The final estimates are published about eight weeks after the reference period, based on about 87% of sample receipt.

An additional shortcoming of mail shuttle collection is that there is little or no contact with the respondent. As a result, respondent rapport is not established and questions concerning data preparation often go unanswered and reporting may not be as timely. Advances in survey techniques provided the impetus to search for a collection methodology that

addresses these sources of response and nonresponse error.

3. CATI BACKGROUND

Beginning in 1984, BLS began to explore alternative collection methods for the CES Survey. The primary objective of this research was to improve response rates. A secondary objective was to improve the quality of the data by identifying and correcting other sources of error.

The earliest test involved soliciting a small number of new establishments into the CES Survey via personal visit. Units agreeing to participate in the survey were assigned to mail or telephone collection at random. From this early test, it was determined that establishments were willing to report the data over the telephone. CATI was not used for this test.

In 1985, BLS launched the start of a multi-faceted testing project to assess the feasibility of CATI in the CES Survey. The CATI system used in these tests was developed by the University of California at Berkeley. The Computer Assisted Survey Execution System proved sufficiently flexible to meet the needs of several research projects using computer assisted techniques. Collection was performed using personal computers.

Approximately 370 CES reporters in Maine and Florida were asked to convert from mail to CATI collection (Werking, Tupek, Ponikowski, and Rosen, 1986). The sample selected represented the range of industries and employment size classes in the overall CES Survey. Over 95% of the units agreed to convert to CATI. Response rates for the earliest publication of estimates improved from 45 to 85 percentage points for these units. Attrition among the CATI units was extremely low. The number of calls required to collect the data was manageable (1.5 calls), as was the length of time required on the telephone (5 minutes). However, the limited sample size prevented meaningful assessments by such variables as industry, size of firm, or payroll type.

4. CATI PRODUCTION TESTS

CATI research has followed a careful plan to test possible effects of the change in collection method on factors such as respondent acceptance, costs, organizational impact, and the quality of microdata. The results of the first small scale tests of telephone and CATI offered the promise of high response rates, acceptable interview times, and relatively few calls per respondent. However, more extensive tests were necessary to fully evaluate CATI and design an

implementation strategy. The CATI production tests were designed to assess program performance, costs, and the effect of change to the organization in a Federal/State environment in which each state is a separate, autonomous unit subject to its own laws and regulations. There are no Federal employees on site. Oversight is conducted from 8 Regional Offices.

As the tests progressed, survey procedures and instrumentation were modified in response to test results and user feedback. This will lead to another level of testing designed to integrate several collection methods into the ongoing CES Survey. Possible implementation strategies include limited use of CATI to bolster weak estimating cells, targeted use for late reporters, or CATI for virtually all respondents.

4.1 Approach

The overall goal of the CATI production tests was to provide sufficient information to develop a cost effective implementation strategy evaluating program performance, costs, and organizational impact across a variety of operating environments and sample sizes. Nine states were initially selected and two states were added after one year.

The nine states were divided into two categories, those with representative samples and those targeting late reporters. The samples in six states (Alaska, Iowa, Maine, Missouri, Nebraska, and Vermont) were representative of the existing state samples. Alaska, Maine, and Vermont were small enough to convert half their mail sample to CATI collection, maximizing the response to the organizational challenge represented by the change in collection method. In three states (Alabama, California, and Florida), the samples targeted consistently late reporters in industries with nonresponse problems and tested the limits of CATI for improving the performance of delinquent reporters.

The CATI production test sample of about 3300 units was initiated between January and June of 1987. In early 1988, further expansion to two more states (Georgia and Mississippi) and 2500 more units over the 11 states was conducted to enhance measures of interviewer productivity, staff utilization, and costs.

4.2 Conversion Procedures

A two step procedure was followed in converting respondents from mail to CATI. First, respondents were asked to verify contact information through the regular monthly mailing. This ensured an accurate contact name and telephone number for the

initial phone call. Next, the mail shuttle form was removed from circulation. In its place, the respondent was sent a new CATI collection form and a cover letter explaining the change in collection method and establishing the date for the first call. Since the data are traditionally obtained from payroll records, the first month call was used to verify the length of the firm's payroll period, as well as collect data and schedule the call for the next month. As part of the evaluation effort, a wide variety of collection events was captured in the data files during the first and subsequent months.

5. RESULTS

To assess the effectiveness of CATI, the following measures were evaluated: timeliness of response, number of calls to collect data, reasons for callbacks, interview time, and attrition. Analysis of these performance measures generally revealed similar results to those obtained under the feasibility test. The following overall results were achieved:

- 85% of the units provided data by first closing;
- 1.7 calls were required to complete a case;
- 59% reported data on the first call;
- 2.5 minutes average interview time;
- 4.4 minutes for the average time the interviewer spent with each case; and
- attrition was about half that of mail.

As might be expected, there was variance in these averages among individual states. This was especially true for the first closing measure and for average interview time. In contrast, the proportion of units reporting data on the first call was remarkably similar.

5.1 Timeliness of Response

The most important aspect of CATI collection for the CES Survey is its ability to speed the collection process. Because appointments are scheduled in advance with each respondent, data are captured at the earliest possible time. Scheduled interview dates are checked by the CATI software against collection deadlines. For example, if the suggested contact date is after the collection deadline for initial estimates, the CATI software prompts the interviewer to request an earlier date. By contrast, in the mail shuttle system, reporting is completely at the respondent's discretion. There is no guarantee that the respondent will complete and mail the form at the earliest possible time. Additionally, there is a 1-3 day delay during mail delivery and the subsequent time needed to key

enter the data once received. Under mail collection, followup of late reporters is generally not conducted until after initial estimates have been produced. Thus, CATI was expected to have a major advantage in timeliness of response.

Figure 1 shows that CATI collection greatly increases the proportion of units reporting data by first closing, from about 50% under mail collection to 85% under CATI collection. However, response rates with CATI varied by state, ranging from 75-95%. Second closing response rates are similarly improved, with 98-100% of the CATI sample data collected, versus 75% by mail. The appropriate application of CATI in the CES Survey could eliminate third closing, allowing the release of final estimates three weeks earlier.

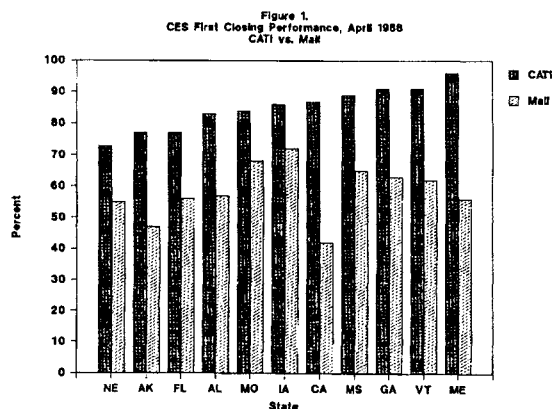
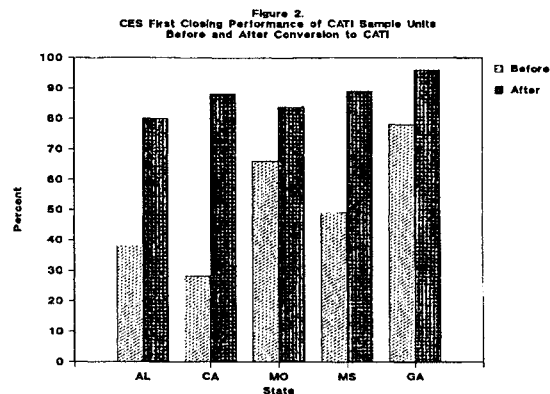


Figure 2 presents a comparison of reporting performance for specific respondents before conversion to CATI with that during CATI. Sample response rates improved by 20-40 percentage points, with Alabama and California showing the greatest increase. A sizable improvement was expected in these states since CATI samples specifically targeted late reporters.

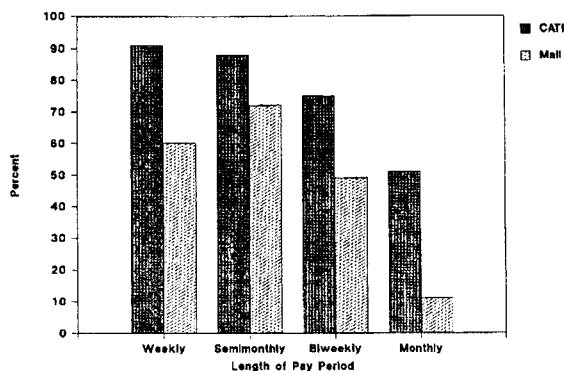


Not only is timeliness critical to the CES, but consistent timeliness is important. The CES estimator links current month's data to that for the previous month in order to calculate an estimate of over-the-month change. Units which report for first closing in one month, but are late in either the preceding or following month may be unusable for first closing estimates in one, two, or even all three months. Under mail, about 2% of the current month units cannot be used in estimation for first closing because previous month's data are not available or unusable due to unresolved edit failures. The scheduling capabilities of CATI and the demonstrated ability to collect virtually 100% for the second closing cutoff virtually eliminates this source of waste.

Response rates are not only affected by timeliness of reporting, but also by length of the pay period and the size of the establishment. The CES Survey uses the pay period including the 12th of the month as the reference period. There are four primary payroll lengths: weekly, biweekly (every two weeks), semi-monthly (twice a month), and monthly.

To be included in first closing estimates, data must be received in the state agency by the last day of the month. This poses a problem for most units with monthly or biweekly payrolls. Since most monthly payrolls end on the 30th or 31st of the month, only about 10% of these units currently report by first closing under mail collection. About 15-25% of the establishments with biweekly payrolls cannot be collected for first closing because the ending date for a biweekly payroll system will fluctuate from one month to the next, relative to the week of the 12th. Virtually all weekly and semi-monthly pay periods end in time for first closing collection (Utter and Rosen, 1984).

Figure 3. CES Response Rates by Length of Pay Period CATI vs Mail



Thus, the payroll composition of a state's sample is a major determinant of the potential response rate. Figure 3 presents a comparison of first

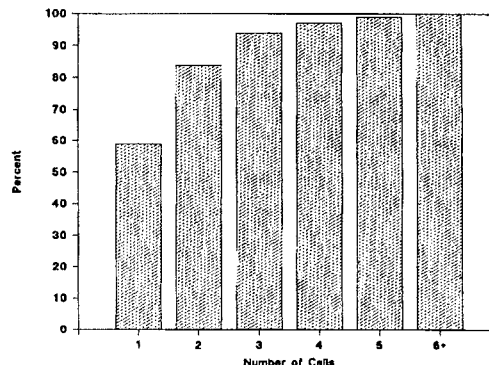
closing response rates for CATI units and mail units by length of pay period. CATI shows a 15 to 40 percentage point improvement over mail.

Size of firm was also a significant factor affecting an establishment's ability to respond by first closing. Large establishments are less likely to report by first closing, due to delays in producing payroll summaries or in the summaries reaching the respondent. When an outside accounting firm prepares payroll summaries, an additional lag is introduced.

5.2 Number of Calls to Collect Data

The CATI interview schedules the date and time for next month's call at the conclusion of the interview. It was hoped that by scheduling such appointments, respondents would be more likely to prepare their data in advance of the next call. Also, the call could be scheduled at the convenience of the respondent. However, if the respondent requests a date that is beyond the first closing date, the CATI software automatically prompts the interviewer to request an earlier date. In this way, both the timeliness objective and respondents' preference are matched.

Figure 4. Cumulative Distribution of CES Cases Completed by Number of Calls



Experience in both the feasibility and production tests indicated that over half the respondents provided data on the first call and, after two calls, over 80% of the data was collected. Figure 4 shows the percentage of cases completed after each call. For the production tests, the average number of calls required to complete a case was 1.7, including no answer or busy signals. State averages ranged from 1.1 to 2.1 calls. About 5% of all calls result in no answer or a busy signal. When these calls are excluded, an average of 1.5 calls is required to collect data.

The two most common reasons for callbacks were because the respondent had not completed the

collection log (40%) and the contact person was not in the office (39%). In only 6% of the callbacks did respondents indicate that data were not yet available, even though some reporters used the initial call as a reminder to compile their data. Furthermore, CATI calls do not appear to disrupt the respondent's normal work schedule, since only 3% of the callbacks were because it was inconvenient to report the data. It is interesting to note that 55% of all callbacks were made the same day as the initial call and an additional 24% were made the following day.

5.3 Interview Time

Two measures of interview length are automatically maintained during the CATI interview: respondent interview time and total case time. Respondent interview time represents the actual time on the phone with the respondent. Total case time includes review time prior to dialing, locating the respondent, actual interview time, and post interview activities, such as entering or changing notes and control file information after hanging up the phone.

Both measures of time have been reduced since the feasibility test and the early months of the production tests. Respondent interview time averaged 2.5 minutes. The primary factor affecting interview time is the number of calls; approximately one-third of the variation in time can be explained by the number of calls made. Total case time averaged 5.6 minutes during the feasibility test, 4.8 minutes during the early part of the production tests, and is currently only 4.4 minutes. This decline can be attributed to four factors: increased interviewer experience; streamlining the CATI interview; automated call dialing; and elimination of programming errors in the software. Thus, early estimates of interviewer costs and productivity were off by 25%.

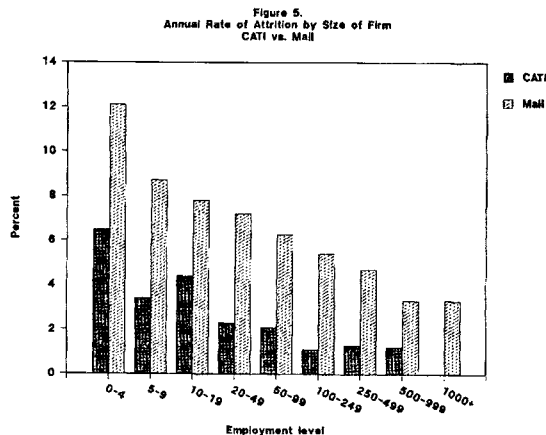
There are indications that the average interview measure overstates the actual time for most respondents, since a few long calls tend to raise the mean. For example, while the average case time was 4.4 minutes, 51% of all cases were completed in 3 minutes or less.

5.4 Attrition

In the CES Survey, sample attrition occurs when establishments go out of business or refuse to continue participation. While the collection method would not affect the out of business component of sample attrition, we would expect collection method to impact the refusal component of CES sample loss.

By comparing the attrition rate of CATI sample units with the rate of overall CES registry cancellations, the effect of mail and CATI collection methods on sample attrition can be shown. The CES registry is the "master file" of valid sample identification numbers. Attrition in the mail sample is captured when cancellations are made against the registry for units that have been inactive for at least six months. CATI sample loss, on the other hand, is captured during the interview and immediately recorded by the collection software. An explanation for sample loss is also obtained.

Data compiled over a six month period from June-November, 1987 indicate that the overall CATI attrition rate was about half that of mail. The annualized attrition rate for CATI units was about 5%, while that of the CES mail survey was about 10%. Attrition by industry was similar — CATI attrition ranged from one-sixth to three-fourths that of mail. Figure 5 shows that results were also similar when attrition rates were examined by size of firm. CATI collection decreases the rate of attrition within each employment size class by one third to three-fourths that of mail. Regardless of collection method, attrition decreases as the size of firm increases. In fact, there has been no sample loss in the CATI collected sample in the largest CES size class (firms with more than 1000 employees).



These comparisons indicate that collection method can make a significant difference in the attrition rate of sample units. The impact of CATI may be attributable to several factors. Most importantly, CATI interviewers establish a rapport with their respondents that reinforces voluntary CES participation. Furthermore, a CATI respondent must overtly refuse to participate, rather than tacitly refusing by simply not responding to the monthly mail shuttle form and postcard reminders. Collection

controlled by the survey personnel rather than the respondent not only improves timeliness but also reduces overall sample nonresponse.

One component of sample attrition that clearly cannot be influenced by collection method is the rate at which sample is lost due to business deaths. The annualized out of business (OOB) rate among CATI units was about 2.5%, comprising around 45% of total CATI sample loss. Because the CATI sample was drawn to be representative of the CES sample, we may infer that the OOB rate is similar for the mail sample. Applying the 2.5% OOB rate to the total number of CES registry cancellations indicates that 20% of the mail sample loss was due to business deaths, leaving 80% of the sample loss as refusals. Thus, the primary source of sample loss is addressable through improved procedures.

The status of the original CATI sample provides a long-term view of attrition under nonmail collection. Table 1 displays the current status of 351 units converted to CATI in 1985 in Florida and Maine for the feasibility test. Over the ensuing 4 years, only 8% of these units no longer respond, versus about 28% of mail units over a similar time period. A total of 195 units have remained on CATI and 103 have been converted to touchtone data collection (TDE), an experimental program where the respondent calls in data using a touchtone phone. It is interesting to note that units which have been returned to mail collection from CATI have continued to report consistently for first closing (about 80%).

Table 1. Status of Original CES CATI Units

Current Status	Units	Percent
Total	351	100
CATI	195	56
TDE	103	29
Mail	23	7
Attrition	30	8

Clearly, CATI is a valuable tool for solidifying respondent participation, offering improvements in sample retention in all industries and employment size classes. These improvements are particularly attractive and cost effective for certainty sample units (generally, those with 500 or more employees), a requirement for accurate estimates. CATI collection of these units not only improves timeliness and data

quality, but also reduces refusals among these irreplaceable sample units.

6. FUTURE RESEARCH

Additional testing is planned in several areas as follows:

Estimation. As late reporters are converted to timely reporters and the sample captured for first closing increases, the estimator will be reviewed to reflect this increased sample.

Scheduling. Procedures for collecting information on data availability in the establishment will further improve scheduling of calls and monitoring of nonresponse.

Attrition. The personal contact factor of CATI appears to significantly improve the relationship with the respondent, resulting in a dramatic reduction in attrition among CATI respondents. Future research will focus on methods to address reasons for nonparticipation. The CATI instrument will be strengthened to include persuasive script to help interviewers effectively respond to negative feedback.

Costs. Cost analyses will be focused on differences in the organizational response to CATI and the productivity of staff conducting both CATI and mail operations. The case study approach will provide a model operational environment for the integration of improved methods with rapidly evolving technology.

Item Nonresponse. There is some evidence that item nonresponse is reduced through CATI collection. Specific questions focusing on this source of nonresponse error will be developed.

ACKNOWLEDGEMENT

The authors would like to acknowledge the contribution of the many members of the Office of Employment and Unemployment Statistics who participated in the development and implementation of the CES CATI Production Test project. Louis Harrell and Wanda Bland participated in the development of all aspects of these projects and assisted in the preparation of materials for this paper. Finally, the authors would like to thank the staff of the 11 State Employment Security Agencies, and the BLS Regional Offices in Atlanta, Boston, Kansas City, and San Francisco for their contributions in testing these developmental methodologies.