

IMPLEMENTING COMPUTER-ASSISTED INTERVIEWING FOR THE CANADIAN LABOUR FORCE SURVEY

Jill Bench, Colleen Clark, Johane Dufour, Ritu Kaushal
Colleen Clark, Statistics Canada, Jean Talon Bldg. 5-A7, Ottawa, Ontario, Canada, K1A 0T6

Key Words: Computer-Assisted Personal Interviewing, Computer-Assisted Interviewing, Data Collection, Data Quality

Computer-Assisted Interviewing (CAI) has been promising survey takers data that is both more timely and of better quality for many years now. The earliest systems were developed for centralized telephone interviewing. The development of systems for personal interviewing followed later as technology for portable computers evolved. However, it is only recently that technology has reached the point where it is cost efficient for large government organizations to adopt computer-assisted personal interviewing as a standard for data collection.

Statistics Canada (STC) recently converted the mode of data collection for the Labour Force Survey (LFS) from paper and pencil interviewing (PAPI) to CAI as the first phase of a multi-year redesign project. Conversion took place over the five months period of November 1993 to March 1994. The LFS is a monthly survey of households that employs both telephone and personal interviewing. Collection with CAI means that each interviewer uses a notebook computer, with the existing questionnaires in an electronic format, for all interviewing. Also, staff who direct and manage interviewers have moved from a paper-based system to one that uses distributed multi-platform computer systems.

This paper gives an overview of how such a fundamental change to the survey's methodology was achieved by describing the collection process before and after adopting CAI, the testing that provided evidence and experience, the preparation of the electronic and human infrastructure, the strategy for converting collection operations, and the impact of conversion on data quality. The paper concludes with some lessons to consider for other organizations contemplating a similar move.

The benefits anticipated from converting the LFS to CAI include:

- i) **Timeliness** - Collection time will be reduced due to electronic transmission and the elimination of data capture. Timeliness is critical for a monthly survey. In the case of the LFS, data are collected for six days with estimates released thirteen days

later. The quality of the final product, the released information, can be improved if the time saved in collection can be used in analysis and in otherwise improving products.

- ii) **Cost** - CAI is expected to improve the efficiency of the LFS by saving data capture costs. After incorporating hardware replacement costs, net savings are anticipated.
- iii) **Data Quality** - An increase in the quality of item data is expected for several reasons. First, data is edited at the time of collection thus allowing invalid data to be corrected at source. Also, some sources of human error are eliminated because of automation. Specifically, branching errors made by the interviewer and error from capturing incorrect data can no longer occur.
- iv) **Generalized Collection Tool For Household Surveys** - Much of the collection infrastructure in place for the LFS is used by other household surveys. Thus, converting the LFS simplifies the path to CAI for these surveys.
- v) **Allows A More Powerful Questionnaire** - Notebook hardware and software can be exploited in redesigning the questionnaire, for example, by employing features such as complicated branching and customized questions. A new questionnaire is currently undergoing development with implementation planned for early 1997.

1. DESIGN OF THE LABOUR FORCE SURVEY

The Labour Force Survey provides monthly estimates of the size, composition, and characteristics of Canada's labour force. Each month approximately 59,000 households are interviewed. The LFS follows a rotating panel sample design based on a multi-stage area frame. Households remain in the sample for six consecutive months. Each month one-sixth of the sample is replaced. Interviewing is done in person in the first month and by telephone in the five subsequent months. Two questionnaires are used. The Form 03 collects data on the composition of the household and the demographic characteristics of household members. The Form 05 asks questions about labour force activity for each person in the

1950
MAY 10 1950
U.S. DEPARTMENT OF AGRICULTURE
WASHINGTON, D.C.

household who is a part of the LFS target population.

2. COLLECTION PROCESS WITH PAPI AND CAI

Collection is carried out by 950 local interviewers across Canada. They are supervised by 100 local senior interviewers. These staff are supported by seven regional offices (ROs) and a division of Head Office (HO) in Ottawa. The working environment for interviewers, senior interviewers, and RO and HO staff involved in collection operations has changed considerably with the introduction of CAI. The collection process required for PAPI employed considerable clerical and data capture resources while the CAI process requires computer-literate resources to support distributed multi-platform systems.

The first step in the LFS collection process begins with selecting the sample on the HO mainframe in Ottawa and transmitting the sample to each RO's mini-computer. Minimal changes to the software were required for CAI.

The next step is to get the sample out to the local interviewers. With PAPI, assignments were printed, collated, and then physically shipped to interviewers using post, courier, or the equivalent. The procedure for CAI is to first download data from the mini-computer to a PC on STC's secure network, where it is encrypted and written to a tape. A person then carries the tape to a communications PC, where assignments are prepared for transmission over telephone lines to interviewers. Physical separation of the secure network PC and the communications PC, which links to public telephone lines, is done to preserve data confidentiality. Interviewers pick up their sample by connecting their notebook computers to a modem and selecting a menu item directing their notebook to call in to the RO. The data are decrypted once they reach the notebook.

The third step is to conduct the interviews and to complete field editing. With PAPI, field editing was manual while with CAI manual editing is no longer required. Edits are built into the notebook so that if invalid data is entered a message is displayed. The interviewer has the option of correcting data or leaving data as it was originally entered. The latter is required since some edits consist of checks on a range of values and, in some cases, values outside the range, although rare, are possible. Also, the interviewer may choose not to pursue an edit failure if she thinks the respondent's co-operation is at risk.

The fourth step in collection is to return data for completed interviews to the RO. With PAPI, interviewers physically shipped packages of

questionnaires to the ROs daily. With CAI, data are transmitted over interviewers' telephone lines to the Communications PC in the RO every evening.

The fifth step is to process data in the ROs. With PAPI, this consisted largely of capturing data from questionnaires. With CAI, data received by the communications PC are written to a tape which is then carried over to the secure network PC, where the data are decrypted. The data are then uploaded to the RO mini-computer.

The final step in collection is to transmit data to the HO in Ottawa where editing, coding, and the remaining processing steps take place. Minimal changes to transmission software were required for CAI.

Machine failures can and do occur. Some of the practices in place to address this include the production of backups at several points. Also, there are persons in each RO dedicated to supporting the notebook computers. Further, notebooks are equipped with removable hard drives for easy replacement. Extra hard drives as well as extra machines are kept in each RO. Data are collected using PAPI as a last resort when time allows.

3. CONVERTING FIELD OPERATIONS

There are four aspects to converting field operations. First, it is necessary to extensively test the new collection mode to investigate feasibility, determine the potential impact on quality, timeliness, cost, and to determine new requirements for field operations. The second phase is to prepare an electronic infrastructure for production. This consists largely of choosing and/or developing the notebook hardware and software. Also, new RO systems are required. The third component of converting field operations is to prepare the human infrastructure. The most costly part of this step is training interviewers and field staff. The last step is to begin using CAI for collection.

3.1 Testing

In 1991 feasibility tests were undertaken to determine what hardware and software could be used for CAI. A major field test, the Data Quality Test, was conducted from July 1992 to January 1993, to assess the impact of using CAI. Specifically, the impact on LFS estimates, data quality and interviewing cost of CAI versus PAPI were estimated. Operational development and evaluation of a notebook computer were also included in this test.

The reliability requirement for the test was to detect a difference of one percentage point in one of the survey's most important estimates; namely, the unemployment rate. A control group of PAPI interviews and a test group of CAI interviews were conducted under similar circumstances to allow a comparison of the two collection modes. The test group and the control group each had a sample size of 3,150 households per month. A total of 6,300 households were interviewed per month with 2,100 households being interviewed each week. Forty-two trained LFS interviewers in four regional offices took part in the test. The test was conducted for 6 months in non-LFS weeks to avoid contamination of LFS data.

Test and control interviewers were paired and assigned to cover the same geographic area. In urban areas, sample interpenetration was done at the city block level. In rural areas, however, the interpenetration was at a higher level due to operational reasons. While it would have been preferable to nest the treatments within interviewers to avoid an interviewer effect and other confounding effects, this option was not undertaken for operational reasons. Rather, in the fourth month, the test group, which had used CAI up until then, used PAPI for one wave of interviews before returning to CAI for the remaining months. Data from the fourth month, when all interviewers used PAPI, was used to estimate systematic non-mode differences that were confounded with the mode effect. These confounding factors were deemed important enough to be incorporated into estimation.

Table 1: CAI Mode Effect on the Labour Force Rates

RATE	STATISTICS	ESTIMATED MODE EFFECT
Employment/ Population	Estimate	-0.10
	Std Dev	0.55
	P-Value	0.86
Unemployment Rate	Estimate	0.08
	Std Dev	0.38
	P-Value	0.83
Participation Rate	Estimate	-0.16
	Std Dev	0.53
	P-Value	0.76

No mode effect was found in the test for the three estimates of rate considered. The unemployment rate, employment to population ratio and the participation rate were not affected significantly by the change in mode. Table 1 contains these results.

The overall mode effect on the nonresponse rate

was estimated to be statistically non-significant at 0.20 percentage points with a standard deviation of 0.36. The mode effect for the *vacancy rate*, which is the ratio of vacant dwellings (unoccupied, under construction, demolished or occupied by out of scope individuals) to dwellings in the sample was estimated to be 0.93 percentage points with a standard deviation of 0.44. Therefore, based on a 95% confidence level, there was a significant mode effect on the vacancy rates. A possible explanation for the mode effect is that interviewers did not have access to a questionnaire after completion and were unable to make corrections if new information became available. *Edit failure rates*, which give the proportion of questionnaires that fail at least one Head Office edit, were somewhat lower for edits that were programmed into CAI but not for others. There appeared to be no apparent pattern by mode for industry and occupation automatic coding rates. Cost estimates based on pay claim data did not appear to be significantly different for the two modes.

The evaluation of the hardware and software used for the test highlighted two problem areas. First, it was determined that the hardware and software combination was not fast enough to carry on an interview that was as smooth as one done with a paper questionnaire. Second, the batteries used (Ni-Cad) were not suitable because they needed to be drained completely before being recharged.

The conclusion from the test was that when implemented, CAI would not have a discernible impact on key series, data quality or interviewing cost. Also, the observed operational problems could be remedied through hardware and software changes. With these recommendations and experience, CAI was implemented in the LFS in November 1993.

3.2 Electronic Infrastructure

The largest part of building the electronic infrastructure to support the CAI collection process was the development of the hardware and software for the notebook computer. The tool implemented for production varied only slightly from what had been used in the 1992 Data Quality Test. The notebook computer being used is an AST PowerExec 3/25SL. Its features include the ability to operate on batteries (Nickel metal hydride) for four hours, a removable hard disk, a floppy disk drive, 14,400 baud modems, 4 MB of RAM, and a 60-120 MB hard drive. The cost of each unit was \$1,800. The notebook software was a combination of commercially available software and software developed internally. There are two

第 10 頁

components to the software; namely the questionnaires and the case management software. The questionnaire software includes skip patterns and edits. Case management software is responsible for moving data to and from the notebook. It also provides tools for the interviewer to organize and manage the work.

Maintaining the confidentiality of data, as required by the Statistics Act, was a critical requirement of the notebook computer and its supporting systems. Several procedures were established to preserve confidentiality. First, physical separation between the secure network PC and the communications PC ensures no electronic connection between networks internal to the ROs and the public telephone lines. Also, data are encrypted whenever they cross platforms. Last, an access control system, with a user ID and password, is built into the notebook computers.

3.3 Training Field Operations Staff

A CAI collection process requires considerably different skills than a paper process. Interviewers need to know how to use their new tool. In the ROs, the need for clerical and data entry resources reduced considerably while there was a requirement to operate the new RO systems and to support the notebooks. These needs were filled by some retraining and by some hiring.

The largest part of training was accomplished by formal training sessions. Sessions were held for five staff groups. Among RO staff, project managers and supervisors in the ROs each received 8.5 days while computer operations and technical support staff received 5 days of training. Among field staff, senior interviewers received 8.5 days of training while interviewers received 7.5 days of training. The total number of training hours for 1,050 interviewers and senior interviewers was 7,975 or 36.25 person years.

Training for interviewers consisted of four modules; namely Home Study, Data Collection, Practice Cases, and Practice Assignment. Trainees were able to familiarize themselves with the basic care and setup of their notebook and modem before beginning classroom sessions. The hardware was shipped to their home for the three-hour Home Study module which also contained a typing tutorial. Trainees then spent three days in classroom training for the Data Collection module which consisted largely of working through scripted interviews. Each notebook was loaded with a number of practice cases that could be worked through independently of classroom training. Trainees spent about 7 hours

going through these cases. Practice cases can be reset thus providing an ongoing training tool on the notebook. The last part of training interviewers was to provide them with a small practice assignment of 15 dwellings. This allowed them to practice their new skills in a real interview situation before using the notebooks for the survey. Three days were allocated for this module.

One part of the training strategy was to have senior interviewers deliver training to their interviewers. This not only saved training costs but gave senior interviewers an opportunity to thoroughly learn the new tool. The last part of the training given to senior interviewers was a one-day session on training skills.

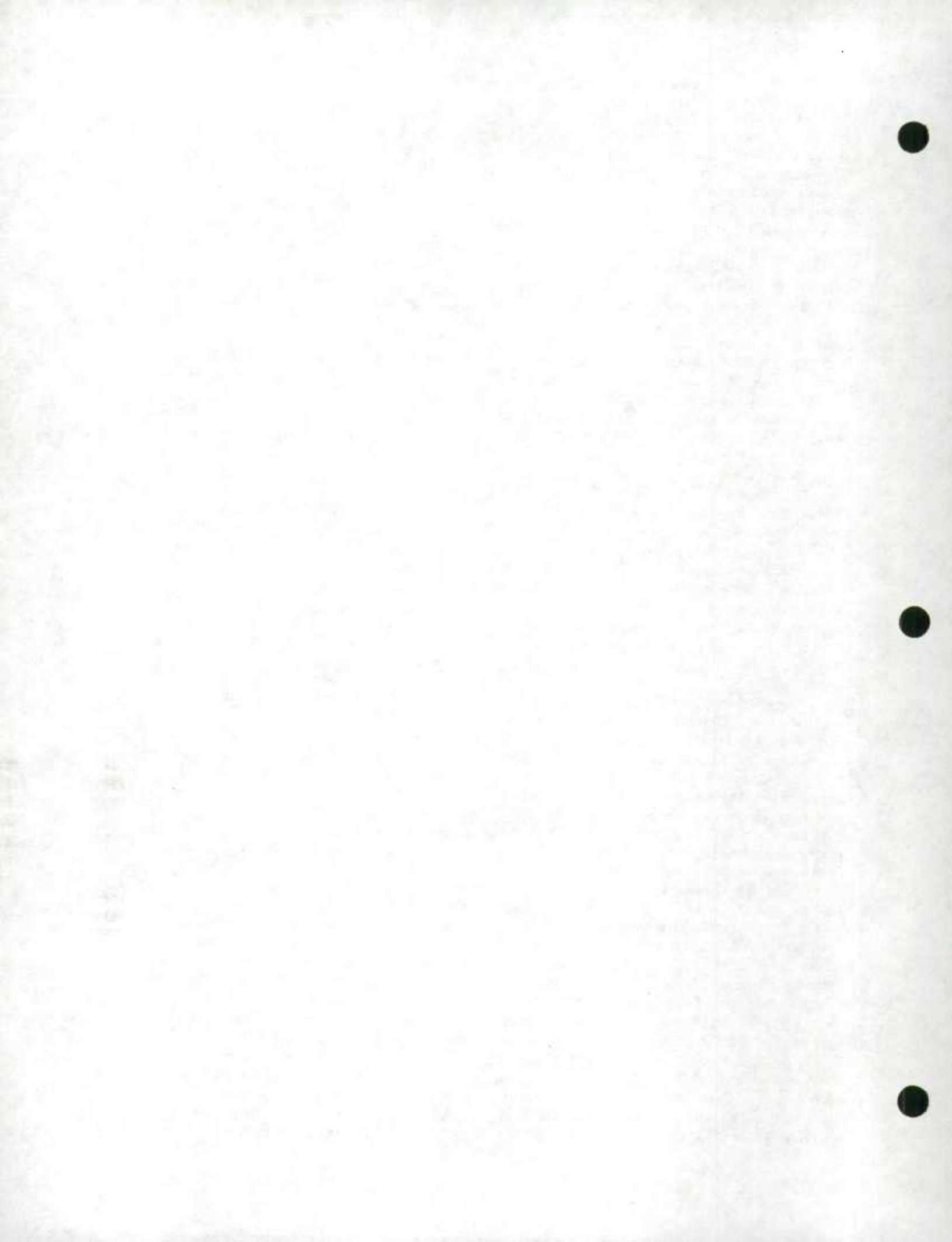
Two impacts of CAI on the job of the senior interviewer beyond the change in collection mode, should be noted. First, senior interviewers will complete a small assignment of about ten dwellings every month in order to practice CAI collection and to keep up to date on software changes. Second, the timeliness and accuracy of information on the progress of each interviewer's collection work has improved. Senior interviewers receive daily reports on their notebooks that give progress as of the previous day of interviewing.

In addition to formal training, the initial surveys done by CAI provided a rich opportunity for on-the-job training especially for RO staff.

3.4 Conversion

Data collection was converted from PAPI to CAI over a period of five months from November 1993 to March 1994. Interviewers were randomly divided into three equal groups. In November, the first group of the interviewers used CAI to complete their assignment. The data collected by these interviewers were used in producing survey estimates. In December, half of the remaining interviewers switched to CAI. No further conversions were done in January and February due to unrelated operational requirements. However, the two-month break provided an opportunity to address problems encountered in November and December, and to learn more about CAI. The final third of interviewers moved to CAI in March.

The factors considered in selecting the phase-in strategy included minimizing the lapsed time between training and beginning to use the new mode for the survey. The potential impact of any learning effect on survey estimates was also a concern. This was achieved by converting gradually over time, by ensuring sufficient training and practice, by randomly



selecting interviewers, and by converting at the same time across the country. Operational complexity and commitments to other surveys were also considered.

PAPI was used as a backup when technical problems occurred with CAI. The extent of this practice was monitored as a measure of the success of conversion. The proportion of dwellings assigned to be interviewed by CAI that were converted to PAPI varied from 2.9% to 6.0%. The maximum was observed in February which was surprising because the CAI interviews in January and February were conducted by the same interviewers. Consequently, it appears that the volume of conversions is not only related to the lack of interviewer experience but to a multitude of other factors.

It is important to note that some of the benefits of CAI are not realized until some time after the initial conversion. The phase-in period was a time of disruption for LFS field operations. This is evidenced by the data quality statistics presented in the next session. There is a significant period of stabilization required before field operations return to the same level of high efficiency as was enjoyed with paper interviewing. It is only then that the anticipated benefits of CAI related to timeliness and data quality are consistently observed. The stabilization period began in March and will likely continue for about a year.

4. IMPACT ON DATA QUALITY

One advantage of gradually converting to CAI is that data quality statistics for dwellings surveyed with CAI can be compared each month with statistics for dwellings using PAPI. Comparisons for four data quality statistics are discussed below. These measures and others related to the quality of the data were examined every month by the LFS Data Quality Committee which consists of quality experts, analysts and operations managers.

The overall *nonresponse rates* were higher for CAI by 1.0 percentage points to as much as 3.4 percentage points. PAPI nonresponse rates from November to February varied from 4.0% to 5.0%. The large difference can be explained by an increase in nonresponse that is due to technical problems rather than to any increase in persons who were not at home, temporarily absent, or who refused to participate. A technical problem, responsible for a considerable portion of the increase, was the loss of data during transmission between the RO communications PC and the interviewers' notebook

computers. Otherwise, CAI had no impact on nonresponse. Despite a concern that the presence of a notebook computer during the personal interview might adversely impact refusal rates, they were stable at 1.2%, regardless of the mode of collection. Furthermore, in some months, the refusal rate was slightly higher for PAPI than for CAI.

For November 1993 to February 1994, the *vacancy rates* for the CAI sample were systematically lower than the PAPI rates ranging between 14.0% and 14.9%. (Note that the 1992 Data Quality Test showed an opposite trend.) The difference, however decreased over the course of the implementation period.

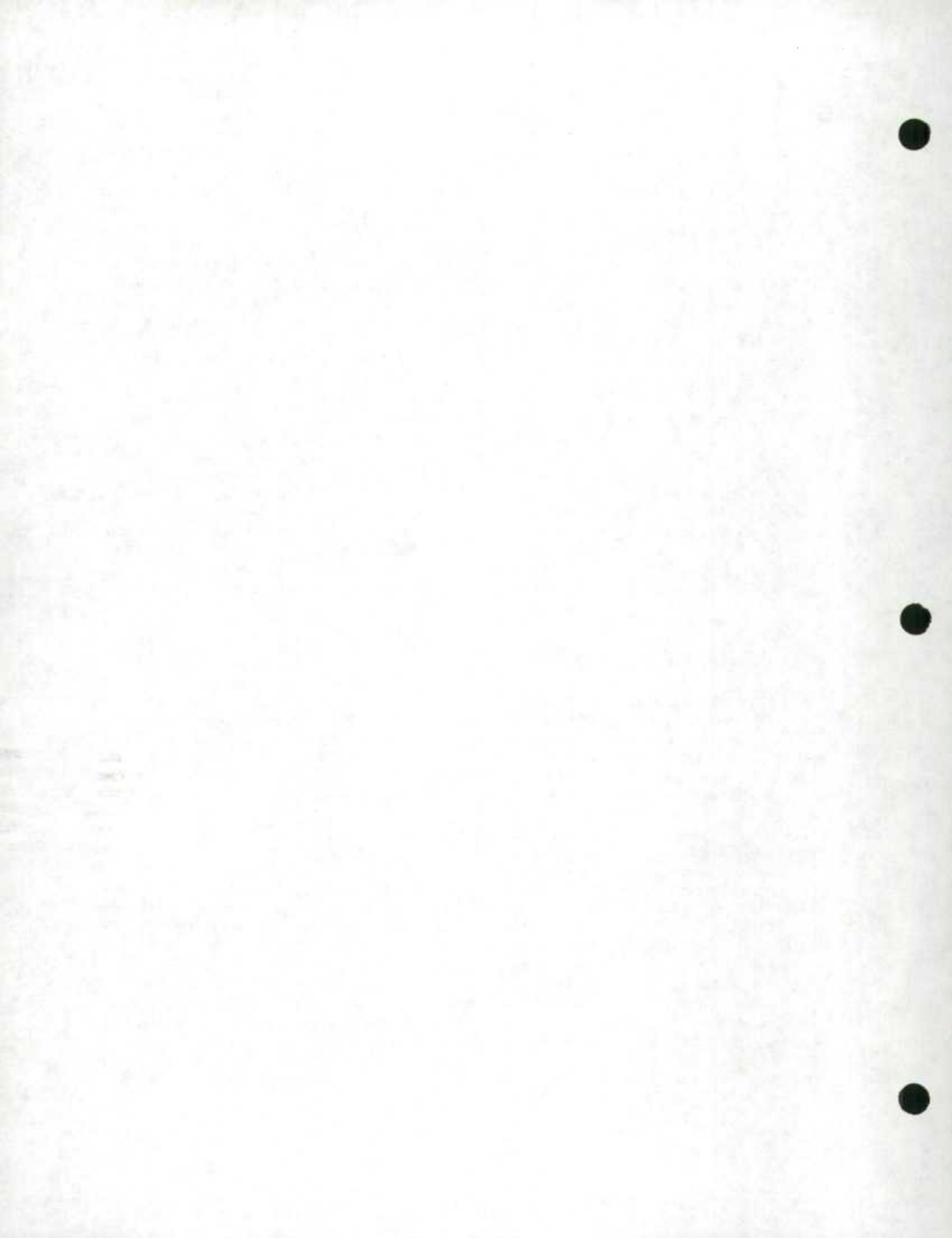
The *interviewer turnover rate* provides the percentage of interviewers who have terminated their employment and who will be replaced. From November to February 1994, this rate varied between 0.7% to 1.1% which is in the usual range. In December 1993, a rate of 0.3% was reached which is the lowest rate for the last twelve months. Hence, the implementation of CAI did not result in any significant interviewer turnover.

One of the anticipated benefits of CAI is an improvement in item quality due to features such as programmed skips and on-line editing. There was some evidence of this occurring from November to February. The *discrepancy rates* provide the percentage of items that failed editing at Head Office. From December to February discrepancy rates were lower for CAI dwellings than for PAPI dwellings. Rates for CAI dwellings ranged from less than 0.1% to 0.1% for the demographic questionnaire (Form 03) and from 0.7% to 1.1% for the labour force activity questionnaire (Form 05). Rates for PAPI dwellings ranged from 0.1% to 0.2% for the demographic questionnaire (Form 03) and from 0.9% to 1.2% for the labour force activity questionnaire (Form 05).

5. LESSONS

Some lessons can be drawn from the Statistics Canada experience for organizations to consider in preparing a strategy for converting field operations to CAI.

The importance of communication cannot be over emphasized. Communication between field staff unfamiliar with computers and computer-literate staff unfamiliar with interviewing is a particular challenge. Keeping everyone abreast of software updates is critical. Involved parties need to be kept informed of plans and progress in a vocabulary that is familiar to



them. A newsletter for interviewers and senior interviewers was created for this purpose. Involving the end-users of the notebook, the interviewers, in the design also proved invaluable.

Strong interviewing skill is the most important characteristic of CAI interviewers. The required computer and keyboarding skills were acquired with ease by LFS interviewers. Interviewer debriefings showed that they felt very positively about CAI and felt more professional using CAI. It is worth noting that there has been no interviewer turnover attributed to converting to CAI.

Conversion to CAI was a period of major disruptions for collection operations. The main sources of these disruptions were technical problems. Consequently, realization of CAI benefits, like timeliness, have been delayed.

Technical problems during transmission of data had a serious impact on data quality as can be seen by an increase in nonresponse. For the conversion period, nonresponse due to technical problems represented 5% to 25% of overall nonresponse. Another consequence of transmission problems was that from November 1993 to the most recent survey no gains in timeliness were realized. There may be some gains in moving data between the interviewer and the RO, but it has taken longer to reach the point of having sufficient data to produce survey estimates. A software rewrite is being undertaken to address problems encountered in transmitting data.

Field experience with the electronic questionnaires has revealed frustrating time lags in moving between components of the questionnaire. A software rewrite is also underway to address this problem. This is expected to reduce the length of the interview, cost and respondent burden.

A significant period of stabilization following the period of conversion is to be expected before the benefits of CAI related to timeliness, reduced cost, and improved data quality are fully realized, and before field operations return to the same level of high efficiency enjoyed with PAPI.

The cost of retraining can be anticipated to be more than the cost of formal training; considerable amount of learning is done on-the-job. Also, the initial surveys with CAI revealed the need to develop new procedures.

To put LFS CAI experience in perspective, 204 PAPI surveys have been done using the same collection methodology and questionnaire, while with CAI only five surveys have been completed since conversion. In a few months, after the stabilization period, the benefits of CAI are expected to be realized.

6. FUTURE PLANS

New software to reduce the length of the interview will be implemented for the September 1994 survey. Reduction in time and cost are anticipated. The rewrite to address transmission problems is scheduled to be used in the November 1994 survey. This is expected to improve nonresponse and timeliness.

STC now has a staff of 950 trained and experienced CAI interviewers who are equipped to undertake other household surveys. The next major development for LFS collection will be the introduction of a new questionnaire early in 1997.

REFERENCES

- Bench, J. , Clark C., Dufour, J., and Kaushal, R. Converting The Labour Force Survey To Computer-Assisted Interviewing. 1994 STC Branch Working Paper (in progress).
- Kaushal, R. and Laniel, N. Computer-Assisted Interviewing Data Quality Test. *Proceedings of the Bureau of the Census 1993 Annual Research Conference*. pp 513-524.
- Singh, M.P., Drew J.D., Gambino J.G. and Mayda F. *Methodology of the Canadian Labour Force Survey 1984-1990*. Statistics Canada Publication. Catalogue number 71-526.

STATISTICS CANADA LIBRARY
BIBLIOTHEQUE STATISTIQUE CANADA



1010217698