

CAI TECHNOLOGIES FOR A CANADIAN LONGITUDINAL SOCIAL SURVEY: RESULTS OF TWO TESTS

Sylvie Michaud and Christel Le Petit, Statistics Canada
Christel Le Petit, Tunney's Pasture, Ottawa, Ontario, K1A 0T6

1. INTRODUCTION

To meet research needs, Statistics Canada will launch a major panel survey of households in 1994. The Survey of Labour and Income Dynamics (SLID) will follow individuals and families for six years, collecting information on their labour market experiences, income and family experiences. A sample of approximately 40,000 dwellings, (yielding about 80,000 individuals) will be selected using an area frame. SLID data will be collected using a decentralized computer-assisted interviewing approach (CAI). Most interviews will be conducted by telephone. Regardless of whether the interview is conducted in the respondent's home or by telephone from the interviewer's home, a lap-top computer will be used to record information rather than paper and pencil. Computer assisted interviewing (CAI) should produce higher quality data, since errors can be reduced using the process. Three tests will be conducted before the implementation of SLID. Two of those tests have already taken place. This paper summarizes the findings of the tests and indicates what is coming next.

2. SLID OVERVIEW AND TESTING STRATEGY

The 40,000 dwellings in the SLID sample will be divided in two panels of 20,000 households each. A panel will be replaced every three years. Respondents will be contacted twice a year. There will be a first contact in January to collect labour information and a second contact in May to collect income data.

Approximately 400 interviewers will conduct interviews, mostly from their homes (telephone interviews will constitute more than 90% of the interviews, the remaining interviews will be personal visits). Each interviewer reports to one of the eight Statistics Canada regional offices who forward the collected data to Ottawa for further processing.

When the tests began, the content of SLID was not yet finalized. However, Statistics Canada conducted two surveys that were used as the base for the content; the Labour Market Activity Survey (LMAS) and the Survey of Consumers Finances (SCF).

SLID is one of the first social surveys at Statistics Canada that will be carried out using CAI. Because the experience with CAI is new, three tests are planned before the implementation of SLID. Two of those tests have taken place so far. The third test will be conducted in early 1993. A major objective of the first test was to verify the feasibility of implementing a survey as large and as complex as the LMAS (the first component of SLID) in a CAI environment. Test 1 was a small scale test designed

to examine the hardware and software required to implement the survey, as well as field operations. The second test (also small scale) tested additional functionality, such as built-in edits, dependent interviewing and electronic transmission. The third test will be a dress rehearsal. It will verify that all operations are functioning, will include do more data quality checks and will be tested with a larger sample. The sample for the third test will be a longitudinal test panel, which will also be used to test some back-end processing and some longitudinal edits and imputation procedures.

The results from the first test changed many of the original plans for the second test, since hardware and software were changed completely between the two tests. The following sections will present the technologies used for each test and the tests' results.

3. TEST 1

3.1 Goals for test 1

The main requirement for the hardware was to have a durable, portable computer that would be easy for interviewers to use, since the interviews conducted using the technology will be a combination of personal and telephone interviews. The first test wanted to identify any physical limitations in the field. Among the aspects rated were the weight of the machine, its durability, the battery life, the screen display, the memory and storage required for collection and back-ups. The first test was also designed to evaluate the turn-around time and the logistics of the customer service support that was required in the event of a machine failure.

A major requirement for the software was that it possess enough flexibility so that requirements from different surveys could be met (each survey will not have its own software or hardware and so it was important to ensure that enough functionality is provided to share the technology among different surveys). Another objective for Test 1 was to evaluate if the length and complexity of the LMAS could be handled by the software. LMAS has a lot of branching and skips in the questionnaire. The test had to ensure that the movement to different questions would be automatic and easy for the interviewers. Also the software had to be able to conduct SLID in both official languages of Canada (English and French).

The first test also determined the interfaces that were required to conduct SLID in the field. Some interfaces were built (like some case management functions) while some others were identified, but left purposely for the second test (like the transmission and the communication/security).

Finally, the first test helped assess the ease of interviewer training and adaptation to the technology. It also produced feedback from the interviewers with respect to the machine performances during interviews (for example was the machine quick enough, was it was "friendly" to use, was the survey easier to conduct on CAI than on paper and pencil. . .)

3.2 Hardware and software for test 1

Several hardware/software options were looked at. The hardware chosen was the DATELLITE 150S. The main reason for this choice was that the DATELLITE 150s is a portable touch screen computer. It does not have an external keyboard. Answers are defined as sensitive areas on the computer; by touching a sensitive area on the computer, the answer is recorded, and the next question appears. The touch screen technology seemed particularly friendly for the interviewers and it was felt it could meet Statistics Canada requirements.

The DATELLITE 150s has the following characteristics:

- Weight: about 5 pounds (without the batteries; closer to 7 pounds with two batteries);
- Size: 2.7" x 10" x 12.7";
- Screen: 9.5" on the diagonal;
- 728 Kb of RAM;
- 20 Mb hard disk;
- It has an application generator called MAGICSLATE;
- It has rubber "boots" on each side of the computer that make it robust in the field;
- It can be plugged into an electrical outlet, or it can operate with one or two batteries. It is possible to recharge the batteries by putting an adapter in a car cigarette-lighter.

The application generator helps build the screens. It was found to be a user-friendly tool. If text had to be entered, an alphabetic keyboard appeared on the screen, and answers were recorded by typing the answers on the keyboard with the pen. A thin strip pad also could appear at the bottom of the screen, to record numerical answers. If many entries had to be made, an external keyboard could be attached to the computer.

The application generator also gives the flexibility to program different types of questions: dichotomous, multiple choice, mark all that apply and scales for rating. A special format for entering dates and hours is also available. A screen may contain only one question or many questions. This is useful in keeping related questions together. The software also allows some pertinent questions to be kept on the top of the screen (for example, the name of the person for whom information is collected).

The DATELLITE 150s did not have enough memory to program the entire LMAS questionnaire in one block. Usually, a block, using the DATELLITE 150s, would contain between 100 to 200 questions, depending on how much previous information had to be carried in the block. The LMAS had to be split into 24 blocks of questions (also called applications) to complete the interview for one person. Movement within an application was relatively easy. One could go back to see previous answers and edits could be done. However, when an application was

completed, it was not possible for the interviewer to go back to that application to change an answer. This was felt to be a major restriction.

Interfaces were built in the programming language C, to help the interviewers manage their case-load. The case management included features such as an initial list of the cases assigned to an interviewer. The list could be shown in two different ways: with the address information, or with information on the best time to call. The interviewer uses the list to select a household to interview. Case management offered the flexibility to assign response codes, either automatically (when a household was completed), or manually (for example to differentiate between different non-response codes). It had an option that let an interviewer suspend a case that was not completed, fix an appointment with the respondent, and bring the interviewer back to the same application where she had left the interview, when the household was contacted again.

Finally, the case management had an option that allowed the language of interview to be changed while the interview was in progress.

3.3 Test 1 methodology

The first test took place in November 1991 with a sample of 140 dwellings. These dwellings were part of a CAI test for two surveys, first the Labour Force Survey and then SLID. The same sample design used in the regular paper and pencil LFS was used for the LFS/SLID first test. The sample was drawn using a multi-stage design, where dwellings are the last stage of sampling. All the persons living in a dwelling form the household. All civilian household members aged 15-69 were eligible for the interview. The dwellings that were found to be vacant, under construction or demolished in the LFS test were excluded from SLID Test 1. This gave an actual sample size of 120 dwellings with household members eligible for interviews.

All interviewers came from one regional office (most interviewers were from the Ottawa region) in order to reduce costs. Seven interviewers aged between 40 and 60 participated in the test. The same interviewers were used for both the LFS and SLID Test 1. They had assignments which varied between 8 to 31 households per assignments. There were on average about 1.8 eligible persons per household.

All interviews were telephone interviews conducted from the interviewer's home, except when a household did not have a telephone number, or if the household had requested to be interviewed in person. This is similar to what will be done in SLID.

There was a three day training session before the start of the test. Procedures were simplified for the first test. For example, tracing respondents that had moved out was not part of Test 1.

The collected information was sent to the regional office using back-up diskettes. Once received in the regional office, the diskettes were then mailed back to the head office, to be further analyzed.

3.4 Test 1 results

Test 1 results were very encouraging. Out of the 120 households in the sample, 116 (96.7%) were in-scope for the survey. Households were considered out-of-scope if all the members from the household had moved out of the dwelling since the LFS test. As mentioned in the previous section, Test 1 did not include the tracing capability, so these households were excluded from the test.

The 116 households interviewed for the first test had a response rate of 92.2 %, which is very similar to the response rates of the LMAS using a paper and pencil interview. Interviews were completed within five working days.

As in a number of tests, battery life was between two and three hours. This created some problems when interviewers were in the field, especially if the batteries were not properly charged. Putting more emphasis on how to charge the batteries solved some of the problems. Interviewers also sometimes asked the respondent for permission to plug their machine into an electrical outlet, if the "battery low" signal was heard during the interview. Focus groups done on respondents from the test showed that they did not mind interviewers plugging the machines in, but they would have preferred the interviewer to have asked for permission at the beginning of the interview, rather than when the battery was low. Since SLID does mainly telephone interviews, the battery limitation was not felt to be a major issue.

The time required to move from one question to another, within an application was adequate. However, it was slow each time an application was completed and a new application was entered. Time was also an issue when the program had to "shell out" to go to a C program (for case management). Interviewers had to fill the time when there was "dead air". Even when it was a problem, it was handled well by the interviewers. However, all the functionality was not implemented. It is felt that if all the functions had been put in place, the time to complete the interview would likely have created problems.

Time was also slow if, within a household, some respondents wanted to be interviewed in English and some in French. Changing the language of interview was slow. The time problems could probably be solved by having a more powerful machine (with more memory, higher speed clock).

As mentioned earlier, the questionnaire had to be divided into 24 applications. Movement was easy within an application, but could not be done between applications. This restricted many of the edits that could have been implemented. No consistency edits were put in Test 1 because of the difficulty of movement. Analysis of the notes given by the interviewers showed problems that were caused because they could not go back in the application.

Test 1 showed that the interviewers enjoyed interviewing with CAI more than with paper and pencil. On average it took longer to conduct an interview using the computer than it did with a paper and pencil interview. However, with paper and pencil, interviewers normally must do a lot of the checking of their documents once the paper

questionnaires are completed. They felt that, on average, they spent the same amount of time, and perhaps a bit less time with the computer, since once a case was closed, it could not be accessed again.

Interviewers also mentioned that it was easier to conduct the LMAS interviews on the computer since all the skips were done automatically.

Based on the results from the first test, because of the limitations of memory and time and the fact that this technology is still expensive, it was felt that at this time, it would not be suitable to use the DATELLITE 150s for a production environment. A second version of the application generator is due to be released, but no definite dates were given. With the second version of the application generator and with a more powerful machine, the decision could be reviewed, because this technology was very well received by interviewers and seemed robust enough to be used in the field. However, this is not likely to happen for a few years.

4. TEST 2

4.1 Goals for test 2

Given the result of the DATELLITE test, another technology had to be found. For the hardware, it was decided to try a regular notebook, since a faster machine with more memory could be bought for less than the touch screen computer cost. After an evaluation, described in section 4.2, the AST PREMIUM EXEC was chosen. Another software was also required. The survey manager from CHRR (Center for Human Resource Research) was selected. The main reason for this choice was that the software had carried a large longitudinal survey in the field and it seemed to offer many of the functionalities that were desired. It also had some post-processing functions that looked interesting.

The second test was conducted in two phases: a test on the labour component and a test on the income component. This gave the opportunity to test the feasibility of turning the data around between two phases of collection to have dependent interviews.

Since it was the first time data on income was collected with CAI, Test 2 was used to identify any operational problems linked with collecting this type of data with the new technology.

Other important new aspects that were introduced in the second test were electronic data transmission and tracing of respondents who had moved.

4.2 Hardware and software for Test 2

For the second test, as mentioned before, an alternative to the touch screen computer was used; a conventional portable computer, 8.5"x11", called a notebook. This technology has been used for many years now by different agencies and is much cheaper than the touch screen technology.

The specifications for the hardware were:

- 386sx;
- Reasonable weight (6/7 pounds);
- 20 Mhz speed;

- DOS 5.0;
- Integrated 3.5" drive;
- VGA screen adjustable for indoor and outdoor work;
- 4 Mb of RAM;
- 60 Mb hard disk;
- Standard keyboard with at least 10 functions keys, no numeric keypad;
- Battery life of at least 3 hours;
- Internal charger;
- No internal modem;

The AST PREMIUM EXEC was chosen as a result of the acquisition process.

Many different software packages have been developed to collect survey data on a standard notebook. The choice of the software that could be used was therefore much wider than when touch screen computers were evaluated. However, the choice of the CHRR application generator seemed very promising, given the functionalities offered by the software for longitudinal surveys (this software is presently used for the "National Longitudinal Survey" in the U.S.). Appealing features of the CHRR software included: the rostering facility; the possibility to slide back to any question in the questionnaire (to correct errors or to review) and then fast forward to where you left off; easy implementation of edit checks; the possibility to quickly change language in the middle of the interview; as well as integrated post-processing.

CHRR software was tested with the DATELLITE 150s touch screen computer. Even though CHRR software could be adapted to work with the touch screen computer, the CHRR software/touch screen marriage was very disappointing.

Some interfaces were programmed in-house. A case management was built to help the interviewers manage their assignments, to link all the sections of the interview, set the final status codes automatically (or manually) and control the flow from section to section. For example, the software was programmed so that an interviewer could not conduct an interview for a person who was not eligible for the survey. In the same way, once an interview was completed for a person, the interviewer could not access that person's information again.

An interface for transmission and confidentiality was also built, which is described in the next section.

4.3 Confidentiality and transmission

At Statistics Canada, very strict rules are followed to protect the confidentiality of collected data. To ensure that the confidentiality was not jeopardized by the new technology, the following procedures were followed for Test 2. These procedures might be different from what will be used in the production environment.

The software PCDACS was loaded on each machine to control the login of the users. Each interviewer had a user name and a password that had to be used to access the survey. PCDACS also encrypted the data written to the hard disk. If the keyboard was not used for 30 minutes, PCDACS asked the interviewer to reenter their user name and password.

Before the interviewers backed up their work, the files were encrypted with PCDACS.

For transmission, the governmental mailing system GEMDES was used. Before the data was transmitted, an envelope was created by PKARC, with a seven digit password. The transmission procedure was entirely automated; the interviewer only had to hook up to an external modem, select the transmit option from the main menu and the system dialled out to GEMDES automatically.

The Regional Office dialled into GEMDES to get the work that had been electronically mailed by the interviewers the day before. The data was then sent to Head Office on a diskette by regular courier.

Back-up procedures were set up in case some problem occurred during the transmission. The procedure was to put a copy of the file on a diskette, which was then mailed to the Regional Office. This turned out to be a good idea.

4.4 Methodology of test 2

The second test was conducted in two phases. The first phase took place in May 1992 and collected labour market activity data. The second phase took place in June and collected consumer's finance data. For this test, a sample of approximately 500 dwellings was selected. The test was conducted in two regional offices, Vancouver and Montréal, with approximately 250 dwellings in each regional office. There were 8 interviewers in each regional office, including one spare interviewer, plus a senior interviewer. The interviewers were all regular LFS interviewers, but they had different computer backgrounds. Some had no typing or computer skills at all while some had computers at home which they used regularly.

The assignments, which averaged 35 dwellings per interviewer, were created using dwelling information from January 1991. No contact name for the household existed. The interviewers had to trace the original people in their assignments if they were not living at the same address. They also had to include new people living with original members of the household. Individuals aged 15 and over were eligible for Test 2.

For the first phase, the interview was divided in four sections. In the contact section we confirmed if the right dwelling/person was contacted. In the membership section we identified who lived in the dwelling. If some members had moved out, tracing information was gathered for those members. If new people moved in, their names were recorded. The demographic section confirmed demographic information for original members and collected the demographic information of new members. Finally, the last section was the survey itself.

Each section was programmed with the CHRR application generator. The LMAS was programmed by CHRR staff, given the tight schedule and the lack of experience of our staff with the CHRR software at the start of the project. The case management system was programmed in-house.

For the second phase, financial information was collected only for people for whom we had labour information.

Therefore, no information was collected for new members of the dwelling, i.e., for people that had moved in since the first phase. Since the time between the two phases was very short, demographic information was not updated during the second phase.

Before the start of the first phase, all the interviewers came to the Regional Office for classroom training, which lasted four days. At that time they were provided with the AST, a modem, the connecting cables, two nicad batteries, an AC adaptor (which is also a charging unit), and some diskettes for backups and manuals.

The interviewers had the opportunity to come to the regional office for their first day of interviewing. Twelve of the fourteen interviewers took that opportunity.

The collection period for the first phase lasted 3 weeks, but interviewers only had two effective weeks to conduct interviews (one week was the regular LFS week). During the collection period the interviewers had to back up their work daily on a diskette. They also transmitted their work to GEMDES every night. Every morning the regional office dialled into GEMDES to get what had been transmitted by the interviewers.

Three days after the last day of collection for the first phase all the data was received at head office, which only had a week to turn the data around before the start of the collection of the second phase. Turning the data around included recreating the assignments and deriving some flags that would drive the interview for the second phase.

For the second phase the classroom training was only one day long. The collection period lasted ten days.

4.5 Some results

Test 2 went very well, but also showed us where we must work to get a better product.

4.5.1 Results of the first phase: Collection of labour data

Out of the 483 dwellings that were in the sample, 85% were in scope for the survey. If all the members had moved out of a dwelling, the new occupants of that dwelling were out of scope, while the original members had to be traced. Since the dwelling information in the assignments was 16 months old, 15% of the people had moved and therefore 15% of the dwellings were out-of-scope.

With the-out-of-scope dwellings removed, the response rate for the first phase of Test 2 was about 80%. This is comparatively lower than the response rate for LMAS with paper and pencil, and the results from Test 1. High non-response seems to be caused by a high percentage of people that could not be traced (8%). People could not be traced because the interviewers did not have a contact name to help them, something that has been shown to be helpful in tracing. Also, the information they were provided with was 16 month old. In production SLID will have a maximum of 8 months between two collection periods.

The refusal rate was high (9%). The main reason for refusal given by respondents was that they had already participated in surveys for Statistics Canada and were not

interested in participating anymore. There was no procedure in place to have senior interviewers do refusal conversions for this test, which usually decreases the refusal rate. Refusals were happening right at the contact portion or very early in the interview, which seems to suggest that refusals are probably more linked to the way respondents were selected than to the use of new technology.

A finding of Test 2 was that the contact section was not flowing properly in every situation. This is a portion of the interview where the interviewers use their interviewing skills in the paper world. When switching to CAI, careful planning must be done to ensure that all possible situations are covered and the system is flexible enough for the interviewers.

The time between each question was very short. However there was still some lag between the sections where the interviewers had to fill some dead air. This should be optimized.

Only one hardware problem was experienced during the collection of the first phase, and no software problems. Both seemed reliable. CHRR software is set up so that each screen has to be initialized by pressing the down arrow. This feature prevented the interviewers from accidentally skipping through more than one screen at a time if they kept their finger on a key for too long. From what was observed during the training and the collection, it is a very good feature, especially for interviewers with no computer experience.

No particular problems were experienced with the batteries. Their life span was about two hours, but less than five percent of the interviews were personal visits and the interviewers managed without problems.

Some problems occurred with the transmission procedure, with two interviewers. For these two interviewers, problems occurred more than once. The back-up procedure for transmission was used and all the data lost was successfully recovered.

This pointed out the need to have a good management system that could keep track of "where cases are located" at any point during the collection period, to help forecast any possible problem.

The average length of contact (i.e., time spent on the phone by the interviewer to conduct all interviews within a household) for the first phase was 21 minutes, which is comparable to the time taken with paper and pencil.

The interviewers generally liked the machine and the software. Some interviewers that had worked with the DATELLITE tested the combination AST/CHRR and were very enthusiastic. They liked the touch screen technology but preferred the functionality offered by the AST notebook. The interviewers who were in the field for Test 2 and who had done the LMAS on paper before liked it much more with CAI.

4.5.2 Results of the second phase: Collection of income data

Dwellings included for the second phase were those dwellings for which we had labour information. This gave

a sample of 403 dwellings, out of which 98% were in-scope. Since the collection of the second phase was done two weeks after the end of the collection of the first phase, tracing was minimal. The overall response rate was 96%, which is higher than the response rate of the regular SCF. This is explained by the fact that all dwellings that were non-respondents to the labour portion had been removed from the interviewers' assignments for the second phase.

For the collection of the income data, the approach of dependant interviews was used. For example, if a respondent reported that they had received some welfare income during the collection of the labour portion, they were probed to give the amount of that income during the income survey. This caused a problem in one case where the flag incorrectly showed that the respondent had reported welfare income. The respondent became angry, since there is a negative connotation associated with welfare. However, dependant interviews reduced the percentage of inconsistent reporting between the two collection periods.

The average length of contacts for the second phase was about 14 minutes. Feedback from the interviewers collecting income on CAI was mixed. Some preferred it while others found it too long and too rigid. Based on their comments we realized that collecting information on income might be done more efficiently with a database approach rather than with the CHRR software.

5. WHAT COMES NEXT

Although the response rate for the first phase of test 2 was lower than the response rate to the LMAS, it did not seem to be caused by the new technology, but rather by response burden and the lack of refusal conversions. Therefore, the plan is still to use CAI for SLID. A third test is planned for 1993, with a first phase in January to collect labour data and a second phase in May to collect income data. This test will be a dress rehearsal, with a sample size of 2000 dwellings.

For Test 3 the new questionnaires for SLID will have been developed. Test 3 will give us the opportunity to test them in the field. The larger sample size will allow us to

look at the quality aspects of the data, by doing some comparisons with some administrative files.

A new contact portion will be developed for Test 3, and more integrated testing procedures will also be tested. New options in the case management should be used, such as transfer of cases from interviewers to senior interviewers and the possibility to send fixes to the software to the interviewers. Test 3 will be our first time in the field with these options.

Also, Test 3 will allow a quantitative evaluation of the effect of feeding back information. There is already evidence, from studies done with the LMA survey, that feeding back some information on the work minimises the seam effect. Test 2 showed that feeding back information can have a negative impact (welfare, for example). However, it is not clear what the effect of feeding back information such as unemployment and pension income will be on the data and on the respondents.

In conclusion, the tests conducted to date with the new technology were very encouraging. The interviewers are positive and the data is ready for processing much faster than with paper and pencil. However, there is still much work to be done especially on the data quality aspect, before we can use the new technology for SLID.

REFERENCES:

- Étude méthodologique des résultats du test 2 de l'enquête sur la dynamique du travail et du revenu, Sylvie Michaud, Christel Le Petit, Julie Girard, Social Survey Methods Division, Statistics Canada.
- Content of the Survey of Labour and Income Dynamics. SLID Research Paper Series: No.1, Statistics Canada.
- Survey of Labour and Income Dynamics; Test 2 report, internal document, Statistics Canada.

ACKNOWLEDGEMENT:

Thanks to Mark Kinack, from Statistics Canada, for reviewing this paper.