
Implementation of CATI Techniques in an Academic Social Science Research Setting

by Dave Odynak¹
and Cliff Kinzel
University of Alberta
Edmonton

Abstract

This paper focuses on the use of computer-assisted techniques for conducting survey interviews by telephone (CATI). A few of the basic and some of the more advanced features of CATI systems are briefly listed. Experiences with integrating the CATI electronic questionnaire into survey planning, training, execution and data processing in an academic research setting accustomed to traditional pencil and paper survey methods are discussed. Special attention is placed on the social aspects of introducing CATI technology into an established research environment.

Experience Meets Technology

Computer assisted telephone interviewing techniques (CATI) have become an extremely important consideration to researchers using the telephone interviewing mode of data collection. The Population Research Laboratory (PRL) has been actively engaged in survey research since 1973 and has administered numerous survey projects through its facilities. These survey projects have involved face-to-face, telephone, and mail-out questionnaires.

At the time of planning for a recent (1989) small survey project in the PRL using telephone interviewing there was very limited use of CATI in social science research centres in Canada. This perceived lack of knowledge and expertise meant that the integration of a CATI system into a research setting such as the PRL would ultimately rely on trial and error. Moreover, the CATI system would be considered as an extension of, rather than a replacement for the traditional mode of conducting surveys. After receiving funding, hiring a researcher for the survey, and then conducting a comparison of existing CATI systems for the microcomputer, the Population Research Laboratory decided to use a CATI system. The main use of the system would be for a small academic survey of public opinion in a localized area (province) rather than for a large national survey or commercial marketing application.

Surveys in an Academic Setting

The Population Research Laboratory functions as the research wing of the Department of Sociology at the University of Alberta. Because the PRL is located within an academic social science setting it must perform

service roles within the University and the Department of Sociology and still run survey projects in a cost efficient manner. The introduction of new survey technologies into academic research settings such as the PRL is necessary to keep pace with the competition for research dollars from non-profit and commercial consulting organizations.

On the one hand, some flexibility is allowed within an academic setting for research into new survey methodologies. Clearly, surveys are conducted in an academic setting for purposes other than trying to make money. An examination of improvements in survey methodologies is encouraged and rewarded in the form of publications, research papers, and graduate theses. On the other hand, survey research can become overburdened with the numerous operating constraints that exist in large bureaucratic institutions like a university. For instance, survey projects are subject to university overhead costs, contract requirements, and public tender on goods. Moreover, conducting surveys in an academic setting can compete for goods and services required for other functions that a research centre like the PRL performs in a university. For example, in addition to providing research facilities for survey projects, the PRL also publishes an academic journal in population, provides staff reprints and discussion papers, consults, coordinates conferences and workshops for the Department and University, and provides staff and supervision of the Department reading room.

Pencil and Paper Techniques

For most surveys, the PRL staff rely heavily on pencil and paper techniques with the majority of the data processing done on the mainframe. Over time, technological improvements are slowly integrated into the survey process. Two examples of bringing new ideas into the survey process are: data entry done directly onto the mainframe replacing data keypunched on cards, and the use of optical scan sheets to collect household information at the start of the interview. However, the advent of CATI in the PRL meant substantial changes to the survey process in a relatively short time-span. This is an important consideration for survey research centres that have well-tested and established procedures for doing surveys.

The administration of the telephone questionnaire was closely modelled after the face-to-face procedures with some modifications. Both types of surveys are characterized by large stacks of paper questionnaires moving through the various stages of survey processing.

The stages in the PRL telephone survey process are described briefly in the following steps. Samples of respondents are drawn using probability sampling techniques (Kinzel 1989). Interviewers and supervisors are hired and trained. Questionnaires are pretested, modifications made and field work commenced. A modified probability sample emerges through random digit dialing techniques and quota sampling at the household level. The interview begins with a well-tested introduction that is relatively successful in gaining cooperation while maintaining informed consent. After an optical scan sheet with an accounting of household members is filled out, an interviewer flips through the paper questionnaire, follows the instructions, asks the questions and records manually on the paper questionnaire the respondent's answers. The questionnaires completed by the interviewers are then edited by supervisors or research assistants. At this stage some of the out-of-range and inconsistent responses are caught. Call-backs and validation checks by supervisors to the survey respondents are made to confirm and obtain additional information and these adjustments are also recorded on the paper form.

Following the completion of interviewing, which takes around two months, responses on the questionnaires for both close-ended and open-ended questions are coded and transferred onto IBM coding sheets. The codes on the sheets are then keypunched into a computer data file. An SPSSx command file is written to read this data file. At this stage, the data are cleaned again using special programs written in Fortran that flag inconsistencies and wildcodes. Typically, for a survey with 1,000 or more respondents and interviews approximately 40 minutes long, it might take up to three months for the coding, data entry, and cleaning to be completed before the data are ready for analysis.

CATI Capabilities

By contrast to the traditional pencil and paper methods previously described, CATI presents interviewers with an electronic questionnaire with the questions, instructions, and choices shown on a computer monitor and answers typed in via keyboard. Much of the editing for wildcodes and inconsistencies is built into the electronic questionnaire so that the cleaning of the data is done continually throughout the execution of the survey. Moreover, the recording of responses by interviewers represents automatic data entry.

Two types of CATI systems exist for microcomputers:

CATI programs that operate on stand-alone computers and CATI programs that can take advantage of networked computers. CATI systems for either type are available commercially. Nicholls (1988) provides an excellent general introduction to the history, use and capabilities of CATI systems. Some notable features available on the commercial CATI stand-alone version purchased by the PRL are:

- o automatic generation of skip patterns or routing
- o handling of special question types including close-ended and open-ended multiple response
- o variable text insertion into questions based on previous responses
- o an electronic coder to code open-ended question responses on-line
- o randomization of questions and question/answer choice
- o arithmetic functions
- o on-line editing of data files
- o export of data into files for statistical analysis

Most of the above features duplicate the traditional pen and pencil questionnaire method using interactive computing to assist interviewers. Even more features are available in CATI systems that are networked. In the networked version of CATI some substantial changes added to the supervisory and administrative functions in the survey process are:

- o automatic monitoring of interviewer performance and survey indicators such as response rate, quotas and completions
- o call disposition monitoring and incidence reports
- o automatic scheduling of call-backs
- o integration of random digit dialing and call scheduling

Research methodologists at the PRL attempted to implement as many of the features of the new CATI

system as possible. The main reason for the thrust was an interest in future use of CATI technology for the telephone component of the PRL's annual omnibus-style survey. It was a major challenge for the PRL to adapt the CATI system to capitalize on the survey administration experience of the facilities' personnel and well-tested procedures for doing surveys. Another challenge was to use the available features of the CATI system to accommodate the many different question types and options that appear on a typical PRL survey.

Survey Planning

The planning of any survey can be an arduous task involving a multitude of details not readily apparent to a layperson. The survey project on public attitudes toward discharged psychiatric patients was no exception. Three separate sub-samples across the province of Alberta were drawn and contacted using random digit dialing techniques and then interviewed. Many of the questions were replicated from the 1986 Winnipeg Area Study which had employed face-to-face techniques (Currie 1986).

Most survey tasks were envisioned and listed before the purchase of the CATI software. After the software arrived several additional steps were added to the planning list but there was still a great deal to learn about the CATI system during the actual survey execution. While much has been said in the literature about the technical aspects of CATI, little attention has been paid to the more social aspects of introducing a CATI system to a research environment (Berry and O'Rourke 1988:458). These aspects were found to be a major concern in the PRL hiring, division of labour, training, supervision, and survey execution.

Project Management

The Laboratory's initial experience with CATI suggests that the person managing the project should have a general awareness of the basic microcomputing operating system, the statistical package the data is headed for, and a wordprocessor that produces ASCII files and routines for data cleaning, in addition to experience with survey methodology in an academic social science research setting. It is doubtful that the CATI system used could be successfully implemented in an academic research setting without personnel competent to merge new technologies with established patterns of doing survey research.

Fortunately, the PRL has staff well versed in survey methodology and also familiar with microcomputers and software. Therefore, the PRL decided to use its own personnel to implement the CATI system rather than to contract the computer programming out. A graduate student researcher familiar with the PRL survey process was also hired to operate in-house. This person was given responsibility for the software programming, ques-

tionnaire design and modification, training, some supervision, data processing, and preliminary statistical analysis. Permanent laboratory personnel looked after sampling, budgeting, hiring, and clerical tasks. The researcher managing the project was to consult with the Director and main research technologist on each major step in the survey process. Temporary staff were hired to do the supervision, interviewing, and coding. This particular division of labour was necessitated by the concurrent involvement of the PRL's permanent personnel in other projects. In other words, the implementation of the new CATI system in the planned survey project depended heavily on where one person was responsible for many of the survey's technical tasks, and data processing leaving little time for supervisory tasks such as monitoring interviewer performance and editing questionnaires from the field.

Training

A major consideration of incorporating the CATI system into a survey research setting is the skill level and training of the interviewers. Only one interviewer in the study had experience with CATI, though not the same CATI system as the one used by the PRL. Interviewers available from the PRL pool of interviewers had no experience with computers. An advertisement in the classified sections of the city's two major newspapers for interviewers with typing skills and computer experience mentioned as an asset brought inquiries from several people with a computing science background but no interviewing experience. Clearly at the time of the study, CATI was not used extensively by survey practitioners in the city of Edmonton.

During the initial stages of the survey, materials from previous telephone surveys, most notably the telephone interviewer's manual and handbook, had to be modified for the CATI system. In addition, materials were developed to explain the editing features of the CATI system. Interviewers were then given this material to digest before the pre-test training session. We suggest that the pretest should be conducted using the full CATI system rather than a pencil and paper techniques. Using this modified training strategy both the CATI system and questionnaire content problems can be dealt with simultaneously. After a very brief training session for the pretest many of the problems of introducing the CATI system to the PRL's experienced interviewers surfaced.

During the training for the pretest, interviewers were shown each item on the pretest questionnaire and instructed on the computer procedures associated with each item. After the brief training and introduction to personal computers, interviewers were then requested to do some trial interviews with a friend over the phone and enter the responses on the computer. Later they would begin interviewing the pretest sample. Severe problems

developed with this training strategy. The interviewers did not find the computers to be user-friendly and in some cases felt great apprehension about even touching computers. Unfortunately, the supervisors had not used the computer either. These problems seem to be common to installations considering CATI use for the first time and have been reported elsewhere (Spaeth 1987:22). The researcher in charge of the CATI system had to solve both the technical and social problems associated with first time computer use. It was clear that a different strategy had to be implemented in training the interviewers for the main study.

Some of the features of the pretest training were incorporated into the main training session. Once again a computerized overhead was used to demonstrate and discuss each question in the study. As in pretest interviews did a few practise interviews on supervisors, the PRL researchers and friends. The resulting interviews were scrutinized carefully before continuing with the main survey respondents. However, the main training session was modified substantially to deal more effectively with both the social and technical problems anticipated for interviewers with the survey.

A major apprehension in using the computer for the first time is a fear of the unknown in terms of operating the computer and coping with a vast array of available software. To help allay some of these fears interviewers were told repeatedly that the computer for this particular application was basically a large typewriter and that they did not need wordprocessing, spread sheet, programming, or other computer-type skills to do this survey. The CATI software essentially turns the computer into a large interactive typewriter for the interviewer. Also useful was the provision of hands-on experience with computers and the CATI

system. Mock interviews were conducted in a large computer lab with the CATI questionnaire to be used. Interviewers could punch the keys appropriate to the response and practise some of the editing functions built into the questionnaire. This strategy had a two-fold benefit. First, the mock interviewing session gave interviewers a less intimidating introduction to the computer. Second, the new strategy also helped to familiarize the interviewer with the questionnaire. In the view of the PRL's experience, careful training can overcome the fears and challenges that the use of computers and the CATI system presents, so allowing, in the long run, more emphasis on the quality of the interview.

Interviewers

In the course of the survey, interviewers were able to overcome their initial discomfort with computers and the CATI system and produce quality interviews. Interviewers reported very few technical problems after the first

week of interviewing. Some interviewers even expressed an interest in learning more about the computer. As mentioned previously, one potential outcome of using the CATI system is that it further centralizes the survey process. In the case of the PRL survey a couple of interviewers were permitted to interview from home using their own computers.

One problem that had to be adjusted during the survey was that after the interviewers became very comfortable with the questionnaire there was a tendency to race through the questionnaire. An interview that initially took around half an hour to complete could be completed under 15 minutes towards the end of the survey. Supervisors detected this problem early and instructed interviewers to slow down their pace in asking questions.

Supervisory Tasks

For the most part, traditional methods of manually scheduling calls, callbacks, keeping track of refusals and calculating response rates were employed in the mental health survey. Supervisors were responsible for keeping track of personnel and callsheets, computing daily interviewer and survey tallies as well as validating completed survey questionnaires. Many of these supervisory functions could be automated with a networked version of CATI. Because of the financial and computing resources available for the study a CATI system for stand-alone computers was purchased. Another consideration affecting a more automated supervisory role was that the supervisors were not familiar with the computer, a severe handicap in trying to implement automation of survey tasks. In the future, supervisors involved with the CATI system would need some wordprocessing skills and basic knowledge of the computer hardware and operating system.

Ordinarily, in a pen and paper telephone survey, the supervisors would also be very heavily involved in editing the questionnaires coming back from the field. In the PRL mental health survey they were not. A major bottleneck developed as a result in the simple editing of the completed questionnaires. While the CATI system can provide assistance to the interviewer on some questions, other questions, especially open-ended types are subject to a high degree of subjective interpretation. A separate computer was required at the time of interviewing to edit the questionnaires on-line. After each shift several interviews would have to be edited and the diskettes backed up. Ultimately, this involved over 800 interviews. The majority of the editing was done by the PRL researcher in charge of CATI, since the version of CATI used by the PRL requires completed interviews, fully edited and validated for merging into a database ready for cleaning and summary statistics.

A result of the awkward editing process in this survey was that the validation of surveys by the supervisors was held up and the majority of data processing, merging and coding of open-ended questions was started after the survey interviewing was completed. This is a serious limitation since validation and prompt editing can contribute substantially to overall survey quality by monitoring interviewer's collection of answers and, of course, serve as a deterrent to the "fudging" of interviews. Once again it is anticipated that many of the problems experienced during these phases of the survey would be lessened in the networked version of the CATI system where computer-trained supervisors can interact with the interviewers in the same computer environment.

Hardware

In retrospect, the PRL was not really set up to conduct a CATI survey given the hardware and personnel requirements. All computers at the PRL are in constant use for administrative, clerical and data processing on other projects. The computers used in the study had varying hardware configurations. Moreover, the design of the survey and implementation of the CATI system centralized the survey process in facilities which were already extremely busy. For the study during the day, only two interviewing stations could be in use and at night, after the PRL permanent staff had departed a maximum of four stations could be operable. In essence, the study had to be squeezed in around the day-to-day functioning of the PRL using machines not dedicated to CATI interviewing. Although the PRL staff was very accommodating on the initial CATI study, clearly alternative arrangements are going to have to be made in terms of space and facilities in future CATI projects undertaken by the PRL. A separate location with networked computers dedicated to CATI interviewing and supervising is one recommendation based on the initial CATI survey experience.

A number of problems in implementing the survey steps were anticipated and almost completely different problems were experienced. Interestingly, the PRL decided to run several paper hardcopies of the study questionnaire as a precautionary measure in case "the CATI thing" did not work. One of the hardest challenges in the planning process was keeping track of diskettes and CATI interviews in the same orderly fashion as the stacks of paper questionnaires coming in from the field in the traditional survey process. Call dispositions and interview scheduling still used pencil and paper techniques while the rest of the survey was handled on-line with diskettes, which were backed up after each interviewing shift. The result was two electronic questionnaires for every respondent. In total, there were 823 respondents in three different sub-samples and 64 diskettes to keep track of during the survey.

Across the study, very few technical problems were experienced. Only one interview had to be redone due to hardware failure. Minor problems experienced during the survey included disk drives making noise during the interview, some computer keyboards having an excessive click, and initially some interviewers would place the diskette improperly in the drive. However, by maintaining a continual backup of the interviewing diskettes technical problems that could result in the loss of several interviews were avoided or minimized.

Telephone headsets, however were a continual source of problem. They were used to allow the interviewer to have both hands free while typing. Although several headsets were tested, borrowed and even purchased during the duration of the study, no satisfactory solution was found. For one thing, different telephone lines (local, WATS and FX used to contact the three separate sub-samples in the survey) required different makes of headsets. By the end of the study some interviewers preferred to use a simple headset for the telephone handset that comes with a phone. Newer models of telephone headsets may alleviate many of the problems experienced with headsets in future projects.

Software and Questionnaire

In the version of the CATI system used by the PRL it was found that all the question types could easily be accommodated. The questionnaire included skip patterns, ranges on question choices, variable text insertion, arithmetic checks for consistency between responses, open-ended questions and close-ended questions with an open-ended category for later coding. One time-saving feature of this CATI system was that the questionnaire could be produced on a wordprocessor and then imported into the program. The remaining programming task for the person constructing the CATI questionnaire was to link up instructions and questions from the wordprocessing file with the logic behind the questionnaire in terms of routing, ranges, and question choices. Introductory instructions and a statement asking for the respondent's cooperation were kept separate from the electronic questionnaire. We felt that in the absence of the networked version of CATI this modification to the introduction would improve the flow of the survey. In total, slightly over 100 questions were asked. The electronic questionnaire was very simple in terms of CATI capabilities and very easy to modify.

Data Processing

From the perspective of data cleaning, the CATI system can make a truly remarkable contribution to the survey process. A major advantage of CATI systems is the automatic generation of skip patterns. Another is the programmable restriction on acceptable ranges on question choices. What this means is that if a response is entered outside a permissible range the computer can

alerts the interviewer about his or her error. With automatic skipping or routing of the questionnaire interviewers can be prevented from asking the wrong question. These two programmable features are clear improvements over traditional pen and paper methods where questions and even entire pages of questions can be missed due to complex routing in the questionnaire. Costly callbacks to correct routing errors can be avoided. Besides contributing to reduced field editing of the questionnaire, we found that the resultant data set from a CATI setup required very little additional cleaning.

The process followed by the PRL was to wait until all interviewing and checking was essentially completed before cumulating all the interviews into one file in order to start coding the open-ended questions. An electronic coder was used where coders coded the responses on-line as the responses appeared on a computer monitor. The PRL coding staff found the electronic coder extremely fast compared to the traditional method of determining codes and placing them on IBM coding sheets ready for keypunching manually. However, the more positive social aspects of coding with several people working around a table were lost when a coder was faced with a computer screen in isolation.

After the open-ended questions were coded on-line, the close-ended and open-ended data files were merged and the data exported to SPSS/PC+. The data were cleaned using SPSS/PC+ DATA ENTRY II and then transported via ASCII portable files to the university mainframe for statistical analysis with SPSSx. Smaller runs were done on the microcomputer. The turnover time between completion of interviewing and the production of a machine readable data file was approximately one twelfth of the time required for traditional methods using keypunching to enter the data and extensive cleaning algorithms.

Cost-Efficiency

It is difficult to evaluate the costs of CATI versus traditional pencil and paper given the varying scope and nature of the telephone survey projects going through the PRL. On the one hand, a quick subjective assessment of CATI costs indicates that, considering the initial start up costs in terms of software and managing personnel the study might have been conducted on a cheaper basis using traditional pencil and paper techniques. On the other hand, a more favourable assessment of the cost-effectiveness of CATI might be forthcoming after several CATI studies are completed and the initial start up costs absorbed.

A rough comparison of the estimated budget and final expenses showed that while computing costs in the form of coding, cleaning, and keypunch are substantially reduced with CATI the costs of supervision in the survey

escalated. The rise in supervision costs reflect problems inherent both in the editing process and division of labour. It is anticipated that with the addition of a networked CATI version that supervisors would spend less time completing survey tallies and sorting calls and more time on editing and validation during the survey process. Furthermore, using a networked CATI system the PRL person managing the project might spend less time supervising interviewers and editing questionnaires.

Overall Assessment

The CATI system is a useful addition and enhancement to survey projects that require telephone interviewing. The speed and efficiency that the electronic questionnaire contributes to the data processing is a major contribution to turnover time and cost. A major consideration in the integration of CATI into an academic research setting accustomed to using pencil and paper techniques is the degree of adjustment required. In the case of the PRL experience with CATI the transition was relatively smooth given the constrained computing resources and lack of computer expertise of supervisors and interviewers. Many of the obstacles to successful implementation of CATI were cleared during this initial run. While traditional pencil and paper methods are still used by the Population Research Laboratory for its main annual survey, the addition of networked computers and CATI version capable of automating many of the current supervisory tasks would further establish CATI in the Department as a preferable survey improvement. Other installations with CATI report the advantages of a more automated system including sample selection (Sharp and Palit 1988). According to a recent survey of 42 academic social science survey facilities in the United States and Canada the PRL is now one of two installations in Canada reporting the use of a CATI system (Spaeth 1990).

References

- Berry, S. and D. O'Rourke. 1988. "Administrative designs for centralized telephone survey centers: implications of the transition to CATI." Pp.457-474 in R. Groves, P. Biemer, L. Lysberg, J. Massey, W. Nicholls II and Joseph Waksberg (eds.), *Telephone Survey Methodology*. Toronto: John Wiley and Sons.
- Currie, R. 1986. *Selected Findings From the 1986 Winnipeg Area Study*. Winnipeg Area Study Research Report No. 12. Department of Sociology, The University of Manitoba, Winnipeg, Manitoba.

Kinzel, Cliff. 1989. *All Alberta Area Study 1989 Sampling Report*. Edmonton Area Series Report No. 65. Population Research Laboratory, Department of Sociology, University of Alberta, Edmonton, Alberta.

Nicholls. 1988. "Computer-assisted telephone Interviewing: A general introduction." Pp. 377-385 in R. Groves, P. Biemer, L. Lysberg, J. Massey, W. Nicholls II and Joseph Waksberg (eds.) Telephone Survey Methodology. Toronto: John Wiley and Sons.

Prewitt, Kenneth. 1983."Management of survey organizations." Pp.123-144 in P. Rossi, J. Wright and A. Anderson (eds.) Handbook of Survey Research. New York: Academic Press.

Sharp, H. and C. Palit. 1988. "Sample administration with CATI: The Wisconsin Survey Research Laboratory's System." Journal of Official Statistics 4(4):401-413.

Spaeth, M. 1987. "Cati facilities at survey research organizations." Survey Research 18(3-4): 18-22.

Spaeth, M. 1990. "Cati facilities at academic survey research organizations." Survey Research 21(2): 11-14.

An important factor in the PRL's decision to use CATI for this project was a conference on state-of-the-art telephone survey methodology, including CATI, attended by the PRL's research technologist.

The PRL purchased the CI2 system for stand-alone computers developed by Sawtooth Software. In addition to the CI2 system the CI2 electronic coder was also purchased from Sawtooth Software.

The PRL Alberta Study (including the Edmonton Area Study) is conducted annually throughout the province of Alberta. Respondents in the City of Edmonton are interviewed face-to-face and the rest of the sample across the province is selected using random digit-dialing techniques and interviewed by telephone.

¹ Dave Odynak Researcher Population Research Laboratory Department of Sociology The University of Alberta Edmonton, Alberta and Cliff Kinzel Research Technologist Population Research Laboratory Department of Sociology The University of Alberta Edmonton, Alberta T6G 2H4 ph. (403)-492-2659