

COMPUTER ASSISTED PERSONAL INTERVIEWING: LESSONS FROM EXPERIENCE

Mary Grace Kovar, National Center for Health Statistics
6525 Belcrest Road, Hyattsville, MD 20782

Introduction

For many years, the revolution in computer technology meant that survey statisticians used mainframe computers for data storage and analysis. Desktop computers are changing that. Faster computers and more memory have moved much analysis away from mainframes; CD-ROM and other such devices are moving storage to the desktop. As desktop and, more recently, laptop computers became common, computers began to change data collection. The computerized records of hospitals, businesses, and farms simplify obtaining data on establishments. Computers made it possible to conduct large numbers of interviews with high levels of standardization and quality control over the telephone¹. Now they are being used in households.

The use of computers for data collection is changing so rapidly that there is not yet even a common vocabulary. Shanks identified four components of computer-assisted interviewing: CATI (computer-assisted telephone interviewing, CAPI (computer-assisted personal interviewing; SAQ (self-administered questionnaires); and DDE (direct data entry)². Nicholls added voice technology. A recent report from the Office of Management and Budget broadened the field to include both data collection and transmission, named it "Computer Assisted Survey Information Collection", and discussed options for automated statistical surveys in three areas: CATI, CAPI, and CASI (computer assisted self interviewing)³. The third area includes: PDE (prepared data entry); TDE (touchtone data entry) and; VRE (voice recognition entry). My favorite acronym is PAPI (paper and pencil interviewing).

Although using computers for data collection is a recent addition to the tools of surveys and survey research, it is obvious that their use is growing very rapidly. CATI is now one of the most widely-used methods of collecting survey data and the number of organizations capable of conducting surveys through CATI has grown enormously⁴. The other applications are intriguing. However, this paper is restricted to CAPI and its purpose is to discuss the implementation of CAPI. The experience of one organization will be used for illustration.

Implementing CAPI

There are many things that must be considered as one moves towards using CAPI. Organizations that are already using CATI must decide whether to rescale an existing CATI program or write one designed for laptop computers. Everyone must investigate whether existing software can handle the most-complicated questionnaire; whether the system will also control interviewer assignments; how to handle data transmission, field editing, and translation to a centralized source. They must also investigate whether current hardware can handle the survey (and enable the interviewers to conduct it). Critically important, but often forgotten, is that everyone implementing CAPI must think about the effect on the organization.

Since this paper is based on experience at the National Center for Health Statistics, some background is needed. The experience comes from being first. When CAPI was implemented for a supplement to the National Health Interview Survey (NHIS) in August 1988, that was only the second time that CAPI had been used for a large national survey in the United States. The first was the National Food Consumption Survey, an experience partially documented by Rothschild and Wilson⁵. That was a survey of 9,000 households; the NHIS has 52,000. With little to guide us, we had to learn by experience.

The National Center for Health Statistics (NCHS) is not a primary data collection organization; we did not have an existing CATI system. Therefore, we did not consider modifying one. That had advantages and disadvantages; the advantage was that we could, and did, use the facilities that a PC has but a terminal does not; the disadvantage was that we had to program from scratch.

We examined the existing software, including CASES⁶ and Blaise⁷, and found them to be unsuitable for surveys that include hierarchical files whose length may be known in advance but whose number is not. However, it appeared that CASS⁸, developed at the University of Wisconsin, might work with some modifications. Therefore, the NCHS acquired a site license for CASS, used it for

a feasibility study⁹, and is still using it for two components of the third National Health and Nutrition Examination Survey (NHANES III)¹⁰.

After working closely with the developer of CASS on the feasibility survey, an innovative programmer recognized that off-the-shelf database software could be used to bring a survey on line quickly and efficiently¹¹. He first used DBXL with Quicksilver as a compiler and then Clipper to program other parts of the NHANES III and the AIDS supplements to the NHIS. This approach worked extremely well for the relatively small and free-standing components of NHANES and for small supplements to the NHIS. However, the most recent use - for the Health Promotion Disease Prevention and AIDS supplements to the 1990 National Health Interview Survey, large complex supplements that are almost complete interviews in themselves - taught us a few more lessons. For the first time the interviewers are required to use CAPI; they don't have an option. Over 200 interviewers are using it for interviewing in 800-850 households every week in 1990. The software works extremely well - there have been very few software bugs, and most interviewers like it. There have, however, been some problems. They include resistance from a few interviewers, many complaints about the weight of the hardware, and learning about data control and transmission.

We also examined the existing laptop computers and chose the Gridlite Plus for the 1988 NHIS. As we gain more experience with a large field staff and move towards putting the full NHIS on CAPI, we wish that we could use one of the faster lighter computers now on the market. However, since purchasing 250 laptop computers is too expensive to do it often, we have to continue using them for the next year or so.

Lessons Learned

What have we learned from this experience that we can share so that you will not have to learn the hard way?

The first lesson is that if you are a pioneer you will encounter unforeseen problems; regardless of how much work you have done, there will be problems. There is no way that small studies, usually conducted using interested interviewers or in

controlled situations, will reveal the problems of full field implementation. It is only when a full field staff of interviewers are required to use new technology that some problems surface. For example, the feasibility study for the NHIS showed that the major issue was screen clarity, the weight of the laptop computers was not seen as a major issue. Therefore, the decision on which laptop to purchase was based on the screen with little regard for weight. Full implementation with all of the 200 plus interviewers employed for the NHIS has shown that weight of the laptop is the major reason interviewers don't like CAPI. Interviewers have back and shoulder problems, they have to walk down many streets and climb up many stairs; the 12-pound computer is heavy and awkward to them.

Work in progress by Couper, Groves, and Jacobs¹² shows that we should have been more alert to the issue of the weight. But that work had not been done in 1988 and, even if it had, the computers available today were not available then.

The second lesson was that one should not decide on a CAPI system before laying out precisely what the total CAPI system must accomplish and one should not implement CAPI before deciding what modifications to existing paper-and-pencil (PAPI) procedures will be needed and making certain that they are made. A CAPI system includes more than the software for a questionnaire; it should be able to handle interviewer assignments, data transmission from the interviewer with error checking, field editing, transmission to a central source, integration with existing programs and other systems, and all the other things needed to control data in a survey. At NCHS we implemented CAPI for questionnaires without the rest of the integrated system fully in place. This implementation of the "front-end" without sufficient consideration of the "back-end" has created most of the problems we have encountered.

The third, and perhaps most important lesson, was that implementing CAPI is different and the organization must learn to work differently.

1. Implementing CAPI can have an unsettling effect on an organization. There is no longer a familiar paper trail - people lose the feeling that mistakes can be corrected later or new code categories added. Certain jobs such as data entry, coding, and

mainframe editing either disappear or are reduced - people worry about their jobs.

2. CAPI is seen as restrictive because people want it to work exactly as the paper and pencil questionnaire did. They don't question whether the paper and pencil version was the best possible solution and they don't think how CAPI can free them to do things they couldn't do before.

3. Things that had previously been done serially must now be done in parallel. Much of the coding and editing can be part of the computer program. Therefore, implementing CAPI breaks down organizational barriers. People responsible for the various phases of a survey, such as questionnaire design, coding, and editing must work together regardless of where they are in the organization or in the sequence of activities they followed when using paper and pencil questionnaires mailed to a central office.

4. Because a CAPI questionnaire can include so much of the coding and editing, it requires a lot more work up front than a paper and pencil questionnaire. Therefore, it requires more time and planning at the beginning. With a well-done survey there will be time savings later, but the time must be re-allocated.

5. Training interviewers for CAPI is different. Leading an interviewer step by step through a complex questionnaire isn't necessary - the CAPI program takes care of skip patterns. Helping the interviewer feel comfortable with the hardware and software is much more important - as we found out when we first put CAPI into the field.

6. The developers of the CAPI software should work closely with the people who are actually going to collect the data. Experienced interviewers know what is going to work for them better than software engineers.

Fourth, as we evaluated CAPI software, we learned what about it was important to us.

We decided not to purchase any CAPI software, no matter how good it is on paper, that is not fully supported (full support includes complete documentation, updates, and immediate answers to

questions from users). Some of the best developers are not equipped to help potential users. They have developed the software for their own purposes, they are willing to share, but they do not have the resources to support multiple users. That is not their fault, and they should not be criticized, but we, as potential users, decided that full support was critical.

We decided not to commit to CAPI software with promises to be fulfilled after payment. It should work NOW.

We decided not to commit to CAPI software that requires an annual fee. Others, having made certain of the two prior points and with a one-time survey to get into the field, might make a different decision. The NCHS conducts continuing surveys and the annual fees add up.

Conclusion

The experience of the past year has changed our future direction. We shall not be using database programs for such a large application as the National Health Interview Survey again. They are neither flexible enough nor generalizable. We plan to develop our own software for laptops working carefully with data-collection organizations to make certain that the laptop software integrates smoothly with the larger system¹³. For some surveys done under contract, we may find it more efficient to use the CAPI software of the data collection organization, but we expect to have the knowledge and experience to know what to ask for in the contract. For some simple surveys or one-time surveys, we may be able to use existing software; we shall certainly continue to evaluate it.

We intend to broaden the use of CAPI until data collection for all NCHS surveys is automated. We believe that the potential savings in time from data collection to data analysis are worth it to us. We also believe, although there is little research to support the belief, that the quality of the data will be better.

Others considering CAPI may make other decisions.

They may decide to delay implementing CAPI because laptop technology is changing so fast. I understand - and sympathize with - their delaying

for that reason, but I don't think that they will be able to play catch up too easily.

They may decide to delay because they want to wait until someone else has developed the ultimate CAPI system. That too is understandable, but someone has to move out front. We have decided to move. Those who want to follow are welcome, but they may learn what I have already learned, the leaders are too busy moving to support those who only want to follow. The leaders will share gladly, but they don't have time to support.

References

1. Nicholls, William L. (1989) *The Impact of High Technology on Data Collection*. CATI Research Report No. GEN-1. Bureau of the Census, Computer-Assisted Interviewing Central Planning Committee.
2. Shanks, J. Merrill. (1989) "Information Technology and Survey Research: Where Do We Go From Here?" *Journal of Official Statistics* 5(1), 3-22.
3. Subcommittee on Computer Assisted Survey Information Collection: Federal Committee on Statistical Methodology. (1990) *Computer Assisted Survey Information Collection*. Statistical Policy Working Paper 19. Office of Management and Budget. Washington DC.
4. Sudman, S. and N.N. Bradburn (1987) "The Organizational Growth of Public Opinion in the United States." *Public Opinion Quarterly* 51(4)Part 2, s67-s78.
5. Rothschild, Beth B. and Lucy B. Wilson. *Nationwide Food Consumption Survey 1987: A Landmark Personal Interview Survey Using a Laptop Computer*. Fourth Annual Census Research Conference, pp 347-356.
6. CSM Staff (1989) *User's Guide to CASES; The Computer-assisted Survey Execution System*. University of California, Berkeley, Computer-assisted Survey Methods Program (CSM).
7. Central Bureau of Statistics, Netherlands.
8. Pallit, Charles D. and Dawn. *CASS; Computer Assisted Survey System*. Survey Research Laboratory, University of Wisconsin.
9. National Center for Health Statistics and Bureau of the Census (1988) *Report of the 1987 Automated National Health Interview Survey Feasibility Study*. NCHS Working Paper Series, No. 32. Hyattsville MD.
10. National Center for Health Statistics (1990): *National Health and Nutrition Examination Survey III: Data Collection Forms*. Hyattsville, MD.
11. Biggar, Ronald S. (1988) *The design and development of a computer-assisted questionnaire utilizing the dBASE III language*. Proceedings of the Statistical Computing Section of the American Statistical Association. pp 259-262.
12. Couper, Mick, Robert M. Groves, and Curtis A. Jacobs. *Building Predictive Models of CAPI Acceptance in a Field Interviewing Staff*. Draft for Census Fifth Annual Research Conference, 1989.
13. Gardenier, John S., Kovar, Mary Grace, Shumate, Pat, Adams, Charles. (1990) *Concept design for the NCHS CAPI software system*. *Computer-Assisted Personal Interviewing (CAPI) Development*. NCHS Working Paper Series, No.34. Hyattsville, MD.