From Products to Systems: Addressing the Needs of CAI Surveys at Westat

James E. Smith, Westat Inc., USA

1. Introduction

Computer-assisted interviewing (CAI) and computer-assisted survey information collection (CASIC) have become important buzzwords in survey research. They represent a wide range of methods for survey data collection using microcomputer and telecommunications technologies that replace traditional paper-and-pencil techniques.

Today, the predominant CAI methods are computer-assisted telephone interviewing (CATI) and computer-assisted personal interviewing (CAPI). Each of these methods relies on interviewers (telephone or in-person) who ask questions given by a computerized questionnaire. By entering responses directly into the computer, interviewers eliminate separate data entry operations at home office. In addition, as responses are entered they are checked for errors and inconsistencies, thus avoiding later re-contacting of the respondents. These advantages combined with its ability to simplify the administration of complex questionnaires, have made CAI the method of choice for many kinds of surveys.

Although interviewing with a computerized questionnaire is at the heart of CAPI and CATI surveys, it is not the whole survey process. Every survey practitioner knows of many other essential activities that make up a whole survey project, such as designing and drawing the sample, determining survey content and questionnaire wording, planning and executing data collection logistics, and various post-collection data processing tasks like sample weighting, sampling variance estimation, creating analytical variables and files, and documenting and disseminating the data. Balanced attention is needed to all of these processes for a survey to be successful in its ultimate goal of collecting and delivering useful information to its intended audience.

After giving some background on CAI systems at Westat, I will comment on our CAI systems integration perspective applied to planning, acquiring, and developing CAI capabilities. Two examples from ongoing surveys will illustrate the importance of systems that support non-interviewing work in CAI surveys, showing the need for better product inter-operability in CAI systems.

2. CAI Systems at Westat

Westat Inc. is an employee-owned research corporation conducting statistical research and related services to agencies of the U.S. Government and other clients. Over the past decades, the course we have taken as survey practitioners using computer-assisted survey methods is similar to other survey organizations. Beginning in the late 1970's, as minicomputers became

more available and powerful, we programmed a CATI system on a multi-user VAX system using readily available screen handling and file handling capabilities of that system..

During the early 1980's this first-generation CATI system was replaced by a new generation of CATI software called Cheshire. This software, which includes several components, was designed and developed in-house to meet our growing needs to perform larger and more sophisticated CATI surveys. Cheshire treats a CATI questionnaire in three integrated parts: flow and computation expressed in a PL/1-like questionnaire authoring language, interviewing screens held in a screen library and called up as needed by the questionnaire language, and a data dictionary defining a hierarchical ("rostering") database used by the questionnaire language.

A Cheshire questionnaire is executed under the control of a run-time program that handles common interviewer navigational and operational tasks, such as backing up, breaking off, restarting, producing audit trails, and interfacing with the operating system and hardware for screen handling and disk access. In a CATI survey, a separate call scheduling program delivers cases to interviewers and performs call prioritizing, work allocation among interviewers, call status tracking, and call processing and scheduling functions. The Cheshire system with the call scheduler has been used in a wide range of CATI surveys from small to large, and from simple to very complex in their questionnaire and operational requirements. The Cheshire system has also proven useful for a number of non-CATI applications such as on-line receipt control and computer-assisted data entry (CADE).

During the 1980's, in addition to rapid developments in PC's themselves, other related developments were important for CAI surveys, particularly the evolution of PC networks and the emergence of portable computers. PC networks opened the door for PC's to replace minicomputers as host systems for CATI surveys, and portable computers permitted computer-assisted interviewing to be carried into the field as CAPI. During this time the run-time component of our Cheshire system was converted to work on DOS-based PC's in anticipation of cost-effective laptop computers and CAPI. Today we conduct CATI and CAPI surveys with the Cheshire system on VAX and PC computers, respectively. Our current systems efforts are aimed at the Microsoft Windows environment with network, database server, and telecommunications technologies for CAPI and PC-CATI..

In general, developments of our CAI systems have been driven by requirements for system capabilities that were otherwise unavailable at the time. For example, several key developments in the first Cheshire system were precipitated by survey requirements that were a step ahead of ours or any other CATI system's reach. The practice of targeting our CAI systems development efforts at particular survey needs as they arise has helped to ensure that the right capabilities are developed at the right time, while also giving into the development process strong technical focus and a clear understanding of who the intended users are.

While we find it essential to continue with this approach to CAI systems development, we also recognize that this approach does not completely address all desirable CAI systems changes and developments for the long term. Therefore, we find it important to plan for CAI systems changes and developments in response to specific survey requirements, while at the same time planning for farther reaching changes, including new technologies and new approaches in CAI.

3. Our Systems Interactive Approach

A key perspective in our CAI system planning, acquisition, and development is to conceive of CAI systems as integrated systems comprised of multiple products that work together. In adopting this perspective we have been encouraged by the fact that today's computer software products often adhere to defacto standards, such as the Xbase database format, that allow some level of compatibility between products. At a more sophisticated level, PC operating systems are now supporting and encouraging inter-operability between products through standards such as Object Linking and Embedding (OLE) in Microsoft Windows. Emerging "plug and play" hardware/software standards also reinforce the trend toward product inter-operability. While we are far from seeing universal compatibility and inter-operability across software products, inter-operability is the goal toward which modern software products and operating systems are headed.

Thus, from our survey practitioner's perspective, it is increasingly important for the CAI software we use to perform well at certain CAI tasks, *and* to perform these tasks with a high degree of compatibility and inter-operability with other potential CAI software. There is an instructive analogy in survey statistical data processing. Here the database handling capabilities of database packages and the analytical capabilities of statistical packages are both needed, and different products excel in each domain. While the data processing for some surveys, particularly relatively simple or smaller surveys, might be handled effectively only within a statistical package despite its limited database handling capabilities, or only within a database package despite its limited statistical analysis capabilities, most major survey applications cannot be served in this way. Thus, what has emerged are statistical packages that can process data residing in database systems using, for example, SQL queries executed from the statistical package. Similarly, a single CAI system that tries to be provide a full range of capabilities needed in a major statistical survey, including complex instrument design, instrument execution in a production environment, and post-collection data processing and delivery may end up trying to do too much, and therefore doing too little for many surveys.

4. Example: Specifying a CAPI Instrument

A computer-assisted questionnaire needs to be specified before it can be implemented. Only in the limiting case of a single researcher who sets up his or her own questionnaire directly in a CAI system can a survey avoid a specification process that involves dialogue, communication, and iteration over questionnaire specifications. A large CAI instrument, for example the family medical survey CAPI questionnaire in the National Medical Care Expenditure Survey, may involve over thousand questionnaire items, many of which are complex rosters, grids, or compound items with overlay screens.

As challenging as implementing this questionnaire in a CAI system is, perhaps the greatest challenge is getting the instrument specified to the point that it can be implemented. This process begins with content determination and general design, including chunking the content into manageable sections based on numbers of items, major skip patterns, cognitive

considerations, etc. Specification then proceeds to the item level using standard item types -- yes/no, pick-all, roster, grid, etc. -- wherever possible. Wording is supplied for questions and responses, and response categories are determined. Skips and other flows, like loops, are specified, as are range and consistency checks to be performed during interviewing.

While it may be possible to incorporate all of these types of specifications directly into the authoring language of a CAI system, we are not aware of a CAI system that fully supports the full questionnaire specification process that involves multiple questionnaire authors, numerous reviewers, the circulation and assimilation of comments from multiple sources, and the incremental building of all the necessary specification elements during this iterative process.

Our approach has been less than ideal, but workable. It is to provide questionnaire authoring system separate from the CAI production system but able to integrate with it. By entering specifications into the authoring system, current specifications are maintained in one place (but with network multi-user access). In addition, during the specification entry process, numerous problems, such as incomplete or confusing revisions, overlapping response categories, and impossible skip instructions are detected, the latter automatically by the authoring system. Various hardcopy or electronic document outputs are obtained from the authoring system to facilitate further iterations in the specification process. Among these outputs is a fully formatted "hardcopy questionnaire" that incorporates most of the specifications into a format familiar from hardcopy surveys. As a final specification, the authoring system generates screen definitions, database information, and other information in files that are directly loadable into the CAI system, or are usable in hardcopy form to implement the questionnaire in the CAI system.

This experience suggests to us that the process of specifying a large and complex CAI instrument, benefits from an authoring system that specializes in managing the specification elements and facilitates iteration, communication, and revisions in this process. At this time, it does not seem practical that the same CAI systems which provides run-time support for questionnaire execution in the field could, or would want to, carry all the baggage of front-end questionnaire specification support. Moreover, many nuances of implementing an instrument in a CAI system are extraneous during the specification stage, and should not be forced on the specification team. Therefore, the development of specialized front-end authoring systems, or at least tools, that are inter-operable with CAI systems is desirable.

5. Example: CAPI Data Transmission

The data in a CAPI survey does little good if it doesn't arrive from the field in a timely way. While mailing floppy disks may serve as an emergency backup procedure, CAPI surveys rely heavily on electronic data communications to move data. In many CAPI surveys, such as the multi-round National Medical Care Beneficiary Survey (MCBS) which maintains more than 200 interviewers in the field almost continuously, data transmission involves more than getting completed CAPI interviews to home office. Field status information, case reassignments, questionnaire updates, software updates, and various housekeeping chores are also supported by the data transmission system.

Currently, MCBS and other Westat CAPI surveys use the telephone system as a "wide area network" connection between CAPI field interviewers and home office. (Other value-added networks have some potential here, as does the Internet eventually, but these are not yet cost-effective or sufficiently widely available CAPI surveys.) Some key requirements of our surveys for a system to support data communications through the phone system are ease of use by interviewers, error detection and correction with understandable warning and error messages, complete transaction logging and communications audit trails, programmable rules to accommodate various project needs and changing needs within a project, and sufficiently fast and error-free data transmission.

An earlier CAPI system which we used involved a popular modem communications package with script capabilities. Although operationally acceptable for some time, this system fell short of providing several of the key requirements to the level of a more sophisticated commercial product which is now in use. This product resides on a communications server at home office and operates under the control of project-specific scripts which specify general and interviewer-specific file transfer rules. The client portion of the system resides on interviewer laptop computers in the field and performs automatic dialing, error correction, and retry attempts to reach the home office server. When connected, the client operates under control of the home office server.

In practice, implementing this seemingly straightforward communications process has raised numerous problems concerning the interaction between communications software and modems, idiosyncrasies and surprising incompatibilities between field and home office modems, and wide geographic variations in telephone line quality. While none of these problems have proven insuperable, and none have caused data corruption or loss, they have demonstrated to us that specialized data communications systems which can anticipate and handle these issues are essential. This seems the only way to provide CAPI interviewers with a turn-key communications system that isolates interviewers as much as possible from the details of the communications process.

A systems integration issue that was faced early in the implementation of CAPI communications was the consistency between the look-and-feel and operation of the CAI system and the communications software. The added training and support costs for teaching interviewers about a different user interface in the communication system, along error handling, would have been high, and would have distracted interviewers from their principal task of learning the CAPI questionnaire. Hence, a high level of integration of the communications component with the CAI system was desirable. This has been accomplished in part by invoking the communications software in the standard CAI system menu, and then operating the communications system with pre-stored commands. However, messages and reports from the communication system are seen by the interviewer and must be handled through the communications software interface. Here is a case where some product integration has been forced, but hopefully will become more graceful under newer software and operating system standards for inter-operability.

6. Conclusion

Methods for computer-assisted interviewing (CAI) and systems that support CAI are becoming increasingly important in survey research. While discussions about these methods and systems often focus on interviewing, a CAI survey includes other important processes that must be successfully carried out to design and complete the survey. For major surveys, I suggest that systems which integrate multiple products into an overall CAI system will be preferred.