EFFECTIVE SURVEY AUTOMATION

John Weisberg

Director, Techneos Systems Inc.
13 Woodhill Crescent, Ottawa, ON Canada K1B 3B7

Jay Beaman

Director, Auctor Consulting Associates, Ltd.
465 Andra Ct., Cheyenne, WY 82009

Abstract: Progress in the options for survey data collection and its effective processing continues. This paper focuses on the rapidly evolving capabilities of handheld computers, and their effective exploitation including links to data captured from scanned questionnaires (OMR and barcodes). The paper describes events in Parks Canada that led to the creation of survey software for handhelds. A current commercial Mobile Computer Assisted Personal Interviewing (MCAPI) application that grew out of the work at Parks Canada is used to demonstrate the current state of the art, and future applications are discussed. Basic decision making criteria for selecting a survey data collection method are provided.

Introduction

From the 1960s to 2000 we have experienced a technology revolution. This revolution has been associated with increasing options for cost-effective data collection and analysis (Beaman 1969, Wintrob 1987). As part of this revolution, an integrated framework for data automation was introduced into Parks Canada, yielding participation in discussions on technology and data at professional meetings such as NERR (Parks Canada 1988; Grim and Beaman 1989a and 1989b; Thomson, Cotter and Beaman 1992; Jaro, Stanley and Beaman 1992). Parks Canada was examining various options that produced electronic data such as traffic counters, data capturing cash registers, push-button questionnaires (Cadotte 1979) and "automated kiosks" (Cotter 1991).

As part of the revolution personal computers became "readily" available so that dependency on support groups that controlled access to mainframe computers was broken. Use of modems for transmission of data from field to a location where analysis could be done became feasible. Transportable computers evolved from luggable to laptop, and finally to truly portable handhelds. Software was also progressing, so that executing a complicated survey on a handheld computer became feasible (Sainsbury, Ditch and Hutton 1993; papers in Westlake, Martin, Rigg and Skinner 1999).

Surveys have been an important data collection method for Parks Canada since the 1960s or earlier (see references in Girt 1981, Parks Canada 1983, Canadian Parks Service 1992). In the early seventies Parks Canada began to collect survey data directly rather than relying on contractors. By the mid eighties concerns about the effort that went into planning for surveys, executing them and getting clean data for use in the analysis had risen. Planning for a summer survey could begin one fall, and not until well after the survey planning cycle had begun again the next year would data even be back from keyboarding. Cleaning of data files, analysis and report writing would then commence. With luck results would be presented in January or February, more than a year after a project was planned.

One option to make the collection of survey data more effective became clear. It was reducing the time and labor in going from an interview to clean files for analysis. By the late 1980s changing technology offered several option. Scanners and "image processing" was one. Also, durable and reliable handheld computers were becoming available. Parks Canada's headquarters saw a great potential for handheld computers and purchased two units for Atlantic Region to try. The idea to be tested with these was that by collecting survey data on a handheld, when an interview was over one had clean data ready to be sent by modem to be "added" to an analysis file. In theory analysis could begin during the survey season and next year's planning could be completed soon after the end of the survey. For most surveys ending after Labor Day, in principle, reports could be ready in early fall. Note that while scannable paper questionnaires can offer rapid turnaround as well, scanning is an extra step, the forms may need to be shipped from the survey site, and control of questionnaire navigation and data input values during the interview are not possible.

Discoveries from the experiment

The Atlantic Region handheld experiment basically went well. A lot of effort went into programming the handhelds and much of that was repetitive and tedious. Similar codelists and editing code had to go into different questions and questionnaires to get them properly displayed on a handheld. Furthermore there was variation in coding because of the differences in machines that were being tested. There was typing, retyping, cutting, pasting and formatting.

What became clear was that:

1. Future gains could certainly be made by automation of much of the programming involved in getting questionnaires into handhelds and displaying properly, branching as required and storing responses. But getting questions setup and running so valid responses were captured was only part of the battle. The structure of output data files to be produced by the handhelds had to be programmed and a program had to exist to "collect" data from a particular survey from multiple handhelds and produce a valid aggregate analysis file. Conditional branches based on a person's responses created some tricky details in producing these files. Because output from the handheld was just input to a program to produce analysis files, even though one had defined codes for asking questions, these values had to be linked into output analysis programs so that, at some point, one could create an analysis file (e.g., SPSS or SAS file).

2. Gains could be made by having questions and responses in a data base so that:
   - they could be easily accessed and formatted into program code producing proper display on a handheld;
   - the responses could be incorporate into response editing code; and
   - questions, codes and their description could be used in defining variable labels and value labels in an analysis file.

   Unfortunately, it was often discovered after questionnaires went into the field that they were not perfect. Making adjustments for this required a great deal of careful programming. Not only was it necessary that code for collecting the data be modified, code for output had to be modified to accommodate the inclusion of new variables and the elimination of any removed. This resulted in a problem in "linking together" data produced with different versions of a questionnaire. When questionnaire revisions occur during the course of a survey, major savings can result from programming that performs data consolidation while automatically accommodating all the modifications.
3. Another problem area was managing the data collected, particularly when the same computer was used for more than one questionnaire. When one is getting data from a computer, particularly datasets for different questionnaires, and bringing each together with other data files previously obtained from the same and other handhelds, errors could arise. Some data were overwritten, some were included two or three times. This meant that, where possible, some files from handhelds had to be recovered and properly added to cumulative files. Where extra records were in data they had to be found and removed.
4. Problems could be avoided and gains made by building standards and rigorous integrity controls into data files produced by the handhelds and supporting these in routines that received and aggregated them to produce the analysis files.

Even if the comments above suggest that there were lots of problems, the experiment was judged to be a success. Atlantic Region acquired more handheld computers and used them in numerous surveys. Other regions adopted the technology. Along with acquisitions went a commitment to addressing the areas in which the effectiveness of handheld based surveys could be improved. By 1992 Parks Canada's Western Region had enough handheld survey activity that they took over responsibility for innovation and development. In 1995, with a change in philosophy and management in Parks Canada, staff responsible for the handheld technology and some other survey work in Parks Canada made the decision to leave, in order to push forward with development of the concept on a commercial basis. The software used in the following examples is the result of the efforts by this group.

The fact that handhelds became a key development area and, say, scanning of paper questionnaires was not a thrust, reflects circumstance. In fact, handhelds are being used with paper questionnaires to improve data collection efficiency for particular surveys. Choosing the appropriate cost-effective data collection technology for a given situation should always be a goal (for more information on scannable forms see LoPresti and Maphtali 1996; Acumen Systems 2000; and Principia Products 2000).

Some Examples of Applications

Prior to discussing what has been achieved and what is seen as important, some examples of Mobile Computer Assisted Personal Interviewing (MCAPI) applications will give an insight in the nature and complexity of what can be done.

Yukon Tour Bus Study

In 1999 Parks Canada conducted a survey of tour bus visitors to sites in the Yukon (Weisberg and Thomlinson 1999). The short tour bus season and tight schedule for each tour group meant that rapid data collection was imperative. An easy to use and reliable solution was critical, as the project manager was located over 1,000 miles away, making on-site supervision impractical.

The project utilized handheld computers with built-in bar code scanners, running the Mobile Interviewer™ software from Techneos™ Systems Inc, and OMR (Optical Mark Reading) mailback questionnaires. An interviewer used the handheld to collect information about each tour group from the bus driver. A mailback post-card sized questionnaire was then handed out to each passenger on the bus, after scanning the form's unique bar coded serial number into the interview record. This enabled individual responses to be linked to group information with a very high degree of accuracy. The overall approach substantially decreased respondent burden, as each passenger was only required to provide personal views, without having to describe the entire group's itinerary—information that the driver provided with greater accuracy. The method met the goals of the project manager. Interviewers liked it as well, considering the opportunity to use the handhelds to be the "fun part" of the job.

Liquor Control Board of Ontario (LCBO) Study

In November 1999, the LCBO used the Mobile Interviewer software to conduct in-store research at its new flagship retail location in Toronto. The questionnaire used branching extensively, first screening respondents based on a series of demographic characteristics. Qualifying respondents were then asked a set of 13 attitudinal questions. The software performed a series of calculations based on mean scores derived from previous research, and identified which of five customer types best fit the respondent—all in real time. Based on the customer type, further in-depth research was conducted in the store.

The alternative, using paper questionnaires, would have required the respondent to wait while data was entered into a laptop and the calculations performed. Using the handhelds made the experience a better one for both respondent and interviewer, according to the LCBO's director of customer insights, who said: "It was easy, instant, and looked cool."

The State of the Art

These examples indicate that the state of the art has advanced quickly. Progress has occurred in both hardware and software.

Hardware
Most papers examining Computer Assisted Personal Interviewing (CAPI) have looked at the use of laptop computers (Sainsbury, Ditch and Hutton, 1993; deLeeuw and Nicholls 1996). There are many problems with the use of laptops in the field:

- Unacceptable battery life (e.g., 1-5 hours);
- Easily damaged (by dropping, or from dust and moisture);
- Poor ergonomics (heavy, difficult to use while standing, and tends to come between respondent and interviewer);
- Cost;
- Operating system (reliability, time to start up/shut down).

Parks Canada pioneered the use of ruggedized industrial handhelds for survey research in the late 1980s in order to address many of these problems. Battery life was 8-10 hours, the units were designed for outdoor use, and were smaller and lighter than a laptop (though at about two pounds, still no picnic for the interviewer to carry for eight hours). Cost remained very high however, and because the units were specialized, only a few vendors existed.

In 1993 Apple Computer's Newton MessagePad created the concept of a powerful yet inexpensive mobile computer. The device didn't catch on, but in 1996 Palm Computing released a smaller and less expensive device, and the market for consumer handhelds took off. Today almost every computer, electronics, and office supply store carries Palm devices. The base model weighs 6 ounces, uses common AAA batteries which provide 40-50 hours of run-time, features an exceptionally easy to use and reliable operating system designed for mobile computing—and costs only US$150.

Palm Computing is the clear leader with over 75% market share at present. Palm has also licensed its technology to other manufacturers, including Handspring, Symbol, Sony and TRG. All Palm devices come with built in programs that can be used to manage project-related information, and third party developers have developed an amazing assortment of add-on programs. Hardware accessory manufacturers have produced data acquisition devices including GPS receivers, digital voice recorders, and temperature and altitude sensing modules. These could be combined with survey software to provide new research approaches.

Current handheld hardware is an excellent alternative to standard paper-based interviewing. Future enhancements will come in the areas of wireless connectivity, which will enable functions such as real-time quota control; larger colour screens with adequate battery life, which will allow the incorporation of multimedia in the survey instrument; and more powerful processors, which will make voice recognition the main input method.

Software
Handheld survey software has also been advancing as the hardware improves. Techneos appears to be the only developer with Mobile Computer Assisted Personal Interviewing (MCAPI) software for Palm OS devices. Some companies had developed for Microsoft Windows CE, but changes in the operating system orphaned the products. Products from additional vendors are likely in the future as the popularity of handhelds continues to grow. Other companies offer custom survey programming. This can be an alternative for very large, ongoing surveys that require a custom interface design, but it is expensive, time-consuming and risky when dealing with numerous ad-hoc research projects.

The issues with the use of handhelds for survey data collection that Parks Canada identified led to the development of a meta-data model. This work established the concepts on which Techneos's software is based. Meta-data is "information about information." Normally interview data is stored as a value (e.g., "1") for a particular variable (e.g., "SATISFAC"). A survey system capable of handling meta-data can also store information about the value: "1" means "Very Satisfied" and was recorded in response to the question "How satisfied were you with the quality of papers presented at the symposium?" Information about which questionnaire was used to collect that value, and on which handheld, in what language, and at what time, is also meta-data.

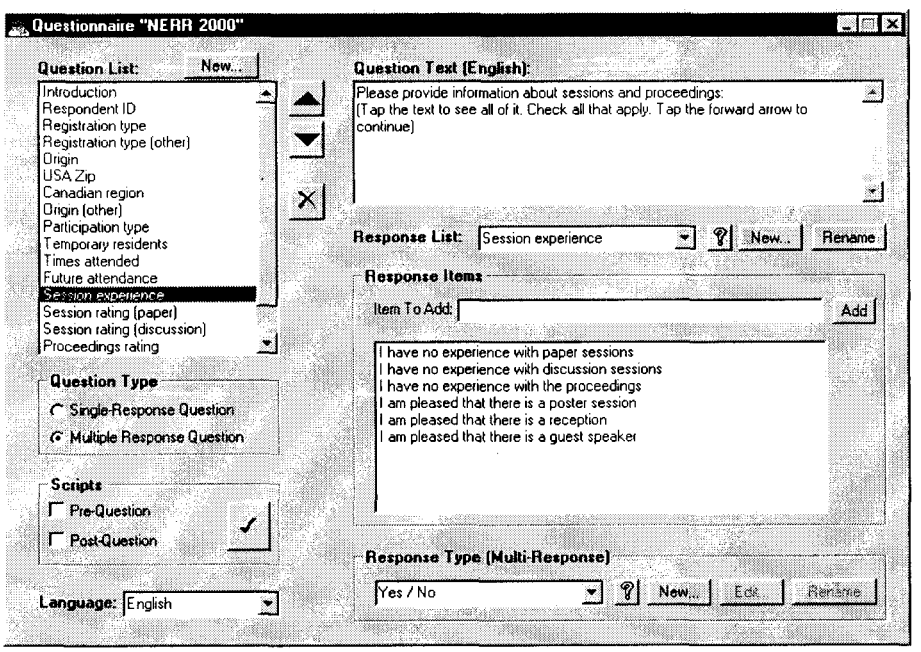
Meta-data is an extremely powerful tool for survey data. Techneos's software is built from the ground up with meta-data in mind, giving it capabilities that are otherwise impossible. Although very few survey software packages currently use meta-data, a new survey interchange standard called Triple-S XML has been developed which will promote its use (Hughes, Jenkins, and Wright 1999). According to the authors, Triple-S XML is "a language for describing survey metadata – that is, data about the survey data: question and response texts, the location of individual data items within the record, valid ranges for responses, and so on." The standard will enable survey designs to be exchanged between any Triple-S XML compliant software packages, eliminating the need for developers to write import and export routines for every other vendor's program. Questionnaires will be more portable, enabling rapidly deployment of the same survey across multiple systems. Triple-S XML is built using XML (eXtensible Markup Language), which is designed to describe what data is—in other words, it is a language for meta-data. XML has applications in many areas, for example to describe the location and possible values for every field on a scannable form. We believe that XML will be rapidly adopted by the survey research industry and will have a profound effect on survey software development.

An MCAPI System

How well does a current state of the art MCAPI system address the issues first identified by Parks Canada a decade ago? We will examine a system consisting of the Techneos Survey Workbench™ for Windows 95/98/NT, and Techneos Mobile Interviewer software running on a Palm OS handheld (all current and most out-of-production models are usable).

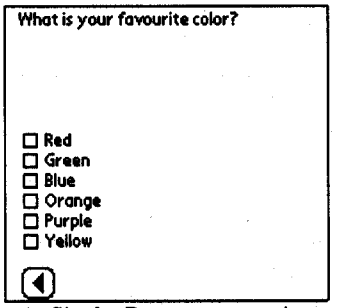
The functions of the Survey Workbench software include questionnaire design, and data communications and management. The program is used to create a questionnaire, download it to one or more handhelds, upload and consolidate response data from the handhelds, and export the response data for analysis.

A researcher uses the graphical user interface to create and name a question, which appears in the Question List. A choice is then made between Single and Multiple Response Question Type—this can be changed at any time with one click. Question Text is added, then a Response Type is selected or a new one created. Nothing more is required to create a questionnaire, unless branching, data integrity checks, or calculations are required. The contention that MCAPI has a high overhead for setup compared to PAPI (deLeewun and Nicholls 96) is no longer valid.

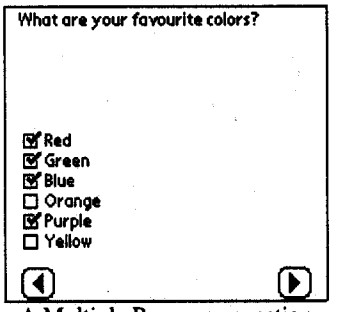


Sophisticated questionnaires can be created using pre- and post-question scripts to initiate actions. Conditional and unconditional branches will vary the path through the questionnaire based on responses obtained. Messages can be conditionally presented to the interviewer. Range and logic checks can be added to improve the quality of data collected, and new values can be calculated on the fly. The questionnaire designer must deal with variables and expressions to use scripts, but no programming is required.

When the questionnaire file is downloaded to the handheld computer, it includes data that describes each question, response set, and script. The Mobile Interviewer software on the handheld interprets this and displays each question in an appropriate manner. The Mobile Interviewer interface is extremely simple, yet the software offers full non-response tracking, the ability to switch interview languages at any time, and automatic recording of information such as respondent number and interview timestamp.



A Single Response question on the handheld. Tapping a checkbox selects that response and advances the survey to the next question



A Multiple Response question allows more than one response to be checked. Responses can be unchecked. After selecting all appropriate responses, the forward arrow is tapped to advance to the next question

Questionnaires may also be printed. The paper version can supplement use of the handheld, or be used instead. Returned forms can be data entered using the handhelds. This is not the fastest data entry technique, but it tends to reduce data entry errors. Once the data is entered, the Survey Workbench software saves time by directly creating a labeled SPSS data set.

The desktop and handheld programs are both very easy to use because of all the work they do behind the scenes. For example, the Survey Workbench software automatically creates and labels the variables required when creating a question, or when switching between single and multiple response. Proper response labels and values are associated with each question. Questions and response items can be deleted and restored at any time just by clicking a button. Entire questionnaires or individual response lists can be copied for use in multiple questionnaires, assuring consistency across an organization's surveys.

A project mode property provides exceptional control over data integrity. In Draft Mode, any changes can be made to the questionnaire, but response data can not be saved on the handheld. Switching to Pre-test Mode activates warnings when certain changes are attempted that can affect data integrity. It also enables modification logging, where every change is recorded in detail. Live mode maintains this high level of data integrity checking, and sets the Mode variable included in every interview record to a different value, so that Pre-test and Live data can be easily separated if required. Each interview record also contains a variable allowing precise identification of the questionnaire version that was used to complete that interview.

The high degree of tracking simplifies the consolidation and export of response data. The software assigns each questionnaire a unique identification number, so response data files from different questionnaires are easily distinguished. Changes such as a switch between single and multiple response for a question are resolved automatically when processing the response data for export. Some export formats allow meta-data such as variable and response labels, and question text, to be included with the response values. This makes it possible to create a ready to use SPSS data set, for example, with just a few mouse clicks.

Cost Comparison with Pencil and Paper

Mobile Computer Assisted Personal Interviewing (MCAPI) competes directly with Pencil and Paper Interviewing (PAPI) in situations where data must be collected on site. The two approaches have opposite cost structures:

	MCAPI	PAPI
Initial cost	**High** - capital investment	**Low** - existing infrastructure
Incremental cost per interview	**Low** - interviewer wages	**High**- interviewer wages plus printing, distribution, data entry, cleaning, data set labeling and formatting

The difference in cost structure means that PAPI is cost-effective for infrequent use and low volumes. MCAPI offers increasing cost efficiencies as the number of projects and interviews rises. The crossover point where MCAPI becomes a better value depends on a number of factors that vary by organization and project. This table shows key items to consider, and indicates which approach is more cost-effective (+). Costs are *fixed* (F) if they do not change based on the number of projects undertaken; *variable* (V) if they vary by project but not by number of interviews; and *incremental* (I) if related to the number of interviews.

Phase and Cost Type	MCAPI	PAPI
System Purchase (F)	- Software, hardware, training	+ Existing infrastructure
Questionnaire Design (V)	+ Reuse elements (response lists, standard questionnaires) + Faster revisions, with tracking + Complex questionnaires possible	- Poor control over standard elements + Faster for very simple, one-time surveys with few revisions - Complex designs unworkable
Printing and Distribution (V, I)	+ No layout, print, storage. Electronic distribution. + Automatic tracking of modifications and distribution + Unlimited inventory at no cost to produce or store + No waste, no delays	- Layout, print, ship, and store forms. Changes cost the same again. - Tracking is difficult and time-consuming - Time and shipping to get forms where needed; lost interview time - Changes produce waste, take time
Interviewing (I)	+ More efficient for complex or long questionnaires + Better response rates; interviewers and respondents enjoy more + Data integrity checks ensure more usable interviews + Automated data capture e.g. date and time stamp + Handles multiple questionnaires	+ More efficient for long open-ended and roster question types - Paper & clipboard is less efficient and attractive - No response checking results in lost due to bad data - Everything must be manually recorded, and done so accurately - Multiple questionnaires difficult
Data Entry Setup (V)	+ Not required	- Must reproduce questionnaire error free; redo if questionnaire changes
Data Entry (I)	+ Not required (done by interviewer) Saves time, avoids introduction of errors	- Additional wages, equipment use. Introduction of errors reduces data quality, may lose some interviews
Data Cleanup (V, I)	+ Minimized by designing questionnaire to ensure entry of clean data + Identify errors during interview and get correct data; more usable records	- Proper data entry not enforced, so varying amounts of cleaning required - Can identify errors, but too late to get good data; lost interviews
Data Set Creation (V)	+ Create formatted, labeled SPSS data set, or export to ASCII (no labels). Saves time, eliminates errors	- Need to reproduce questionnaire design again, error free; redo if questionnaire changes

Some rules of thumb for determining when to use MCAPI instead of paper:

- Use MCAPI for complex questionnaire designs, or when scripting can reduce the interview length for respondents;
- Use paper for many people in a very short time (i.e., to get an adequate sample in the available time, at reasonable cost).

Mobility is the characteristic that distinguishes MCAPI from other data collection methodologies such as phone interviews, mail-out surveys, or web surveys. Mobility is highly desirable in two common situations:

- When the population of interest is most easily reached on site, because they are widely dispersed the rest of the time or have a low incidence in the general population. Park trail users are a good example of this situation.
- When the setting matters. Instead of dealing with anticipation or recollection, you can research an experience while the respondent is in it. Participation tends to be high in this situation. Observational studies are also an excellent option.

It can be seen from this brief overview that the issues with Computer-Assisted Personal Interviewing that have been raised in the past are generally resolved by current Mobile CAPI systems. The use of handhelds with appropriate software is a viable replacement for Pencil and Paper Interviewing (PAPI) in terms of its capabilities; in fact, it is in many ways superior in terms of data quality, time, and cost. However, like every other survey data collection methodology, the suitability of MCAPI must be evaluated based on the overall requirements of each research project.

(A limited number of screenshots were included in the paper to save space. A free evaluation copy of the software can be downloaded from www.techneos.com).

References

Acumen Systems (2000) "AcuForm" http://acumensoft.com/acumen.htm

Beaman J.G.; Leicester J.B. 1969. "Computer Usage in Recreation Survey Techniques". Journal of the Canadian Association of Health Physical Education and Recreation. 35(5):25-28.

Cadotte, E; 1979 "The Push-Button Questionnaire: A New Tool for Measuring Customer Satisfaction" The Cornell H. R. A. Quarterly, Feb.

Canadian Park Service (1992). Socio-Economic Impact Analyis of Parks: Reference List. Ottawa: Parks Canada.

Cotter, M. (1991) Smart Kiosk: Evaluation Issues, Role and Reference Manual. Ottawa: Parks Canada (now only available by request from auctor@igs.net).

deLeeuw, E. and Nicholls, W(2nd). 1996 "Technological Innovations in Data Collection: Acceptance, Data Quality and Costs" *Sociological Research Outline,* 1(4): http://www.socresonline.org.uk/socresonline/1/4/leeuw.html

Girt, H. (1981) Summary Report of the Activities of the Socio-Economic Division 1975-1980. Ottawa: Parks Canada.

Grimm, S. and Beaman, J. (1989a) "The Park Use Related Data System (PURDS) Concept". Visitor Management Conference, January 1989, University of Waterloo. Also in Serving Visitors and Managing Our Resources : Proceedings of a North American Conference on Visitor Management for Parks and Recreational Areas, Robert Graham and Richard Lawrence (Eds.), Environment Canada, Canadian Parks Service; Tourism and Recreation Education Centre, University of Waterloo.

Grimm, S. and Beaman, J. (1989b) "Park Use Related Data Recording: A New Direction for the Canadian Parks Service". Proceedings of the 1989 Northeastern Recreation Research Symposium, April 3-5, 1989, U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station, General Technical Report NE-132, pp. 69-76.

Hughes, Keith; Jenkins, Stephen and Wright, Geoff. 1999. "Triple-S XML: A Standard Within A Standard." Presentation at "ASC 99... leading Survey & Statistical Computing into the new millennium," 22 – 24 September 1999, Edinburgh, Scotland. http://www.asc.org.uk/ See also http://www.triple-s.org/

Jaro, M.; Stanley, R.; and Beaman, J.(1992) "The Anatomy of Complex Data : The underlying Structure of Questionnaires and Forms". Proceedings of the 1992 Northeastern Recreation Research Symposium, U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station, General Technical Report NE-176. April 5-7, 1992, Saratoga Springs, New York.

LoPresti, F. and Maphtali, Z. (1996) "Statisticians' Lib: Using Scanners and OMR Software for Affordable Data Input." Connect, Fall: http://www.nyu.edu/acf/pubs/connect/fall96/SocSciLopScanF96.html

Parks Canada (1983) Reports From Parks Service Socio-Economic Studies. Ottawa: Parks Canada.

Parks Canada (1988) The Park Use Related Data System (PURDS) Data Model: Concept Paper (edited by M. L. Cotter) Ottawa, ON: Environment Canada (now only available electronically from auctor@igs.net).

Principia Products, Inc. (2000) "Remark Office OMR" http://www.principiaproducts.com/office/index.html

Sainsbury, R.; Ditch, J.; and Hutton, S. 1993 "Computer Assisted Interviewing." Social Research Update, 3: http://www.soc.surrey.ac.uk/sru/SRU3.html

Thomson E.; Cotter M. and Beaman, J. 1990. "Effective Management of Parks and Recreation Information". Proceedings of the 1990 Northeastern Recreation Research Symposium, February 25-28, 1990, U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station, General Technical Report NE-145, pp. 103-110.

Weisberg, John and Thomlinson, Eugene. 1999. "Improving Field Data Collection with Mobile Computers." Presentation at "Maximize Your Investment in Market Research" symposium, November 19, 1999, Winnipeg, Manitoba, Canada. Available on request from johnw@techneos.com

Westlake, Andrew; Martin, Jean; Rigg, Malcolm and Skinner, Chris (eds). 1999. "New Methods in Survey Research 1998." Proceedings of the Association for Survey Computing (ASC) international conference, a satellite meeting for Compstat 98, 21–22 August 1998, Chilworth Manor, Southampton, UK http://www.asc.org.uk/

Wintrob, Suzanne, 1987"Computers in Outdoor Education",Computing Canada, September 17, 1987, p. 8-9 (OOFF).