Web Data Collection with Long and Complex Questionnaires

James O'Reilly and Craig Ray
Westat
Rockville, MD
(jimoreilly@westat.com)

Abstract

Using the Blaise system Westat has implemented Web surveys with large numbers of questions, complex routing, and other features beyond those found in most Web surveys. One study is a multi-modal (CATI and Web) household interview that collects nationally representative data on the public's need for, access to, and use of cancer-related information. The second study is a distributed CAPI application where respondents are interviewed in a clinical setting by an interviewer using the Web survey. We will demonstrate these applications and summarize our experiences.

Background

Westat has substantial experience using the Internet and Web for survey management and survey data collection. In the past, most of the Web surveys were done using internally developed systems based on the Microsoft Active Service Pages (ASP) dynamic Web technology and Microsoft SQL Server. This approach allows the developer to tailor pages and the process to a fine level. The highly customizable ASP approach, however, means more challenging programming of the user interface and the routing through the interview for surveys with many questions, skips, and edits.

The release by the Blaise system of its Internet Services (IS) technology offers a number of important features:

o   Strong multi-modal capabilities with the same Blaise language and tools for development of computer-assisted personal interviewing (CAPI), computer-assisted telephone interview (CATI), and computer-assisted Web interviewing (CAWI) surveys.

o   Full implementation of the Blaise rules engine enabling for the Web the same complex routing, edits, and related questionnaire features of Windows applications.

o   A higher level development model for both the question routing and the user-interface elements, lessening the need for specialized programming.

o   Support for large-scale questionnaires of hundreds or even thousands of items.

o   Strong multi-lingual capabilities.

After testing and evaluation, Westat has implemented two Blaise IS Web surveys for studies in which the requirements seemed well-suited to the system's capabilities.

HINTS 2

The Health Interview National Trends Survey, sponsored by the National Cancer Institute, collects nationally representative data about the American public's use of cancer-related information.

For HINTS 2 the sponsor was interested in exploring the efficacy of using a Web version of the questionnaire along with the primary CATI interview. Detailed information on the design, methods and results will be reported in the coming months. For this presentation we focus on the Web survey techniques and experience.

The Web participants were enlisted during the telephone screening process. Subjects were asked if they had a computer, and could connect to the internet. A percentage of those who qualified were randomly assigned to the Web survey group and asked if they were willing to take the survey over the Web. Those who agreed were contacted by email and mail with the survey's URL and their unique username and PIN. Cases that did not complete the Web process within two weeks were transferred to the telephone interviewing process.

HINTS 2 Web Results

Slightly less than half of those that agreed to complete on the web actually followed up and did so. A significant number of those that did not follow-up were later contacted and completed by telephone. Overall, providing the option to complete on the web reduced the overall response rate relative to not providing this option."

As with many Web surveys, the HINTS 2 experience shows the challenge of enlisting public agreement to participate, and the high percentage of those that agree who do not follow through. HINTS 2 might also be described as a particularly difficult Web survey to offer to the public. Subjects must first be contacted by telephone, questioned about survey qualification and internet and email access, and then agree to the Web option. That is followed by an email contact with the URL, username and PIN, and the step of connecting to the survey. Once the subject begins, the typical interview might involve more than 150 questions.

Development

Blaise programmers at Westat took the CATI interview Blaise code and adapted it for Web administration. Much of the work consisted of:

- o recasting question texts from the interview-administered context to that for self-administered
- o defining and applying user-interface settings with the Blaise mode library
- o working with the Blaise IS workshop to adapt the interview XSL stylesheet to implement specialized user-interface elements
- o tailoring stock ASP pages that worked with the Blaise Web processes for user authentication, page handling, and a number of other tasks
- o building the Web case management system

In general, we found the IS system exhibited significant advantages in terms of its robust development process and supporting tools. Programming the instrument, refining the look-and-feel so that it was suitable for everyday Web users, as well as testing, revision, and implementation all met or exceeded our expectations.

User Interface

The HINTS 2 page design is shown below, including

- o Two "tabbed" pages at the top, one for the survey, and one with general help information.
- o Icons for question-level help [ ] and entering comments [ ].
- o Don't Know and Rather Not Answer selections for virtually every question.
- o Previous, Next, and Save and Exit buttons.

Figure 1: HINTS page design
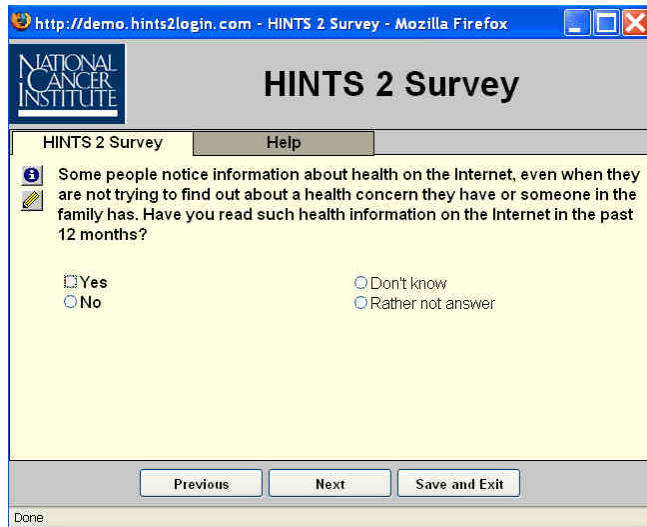
Skipped Questions

An important issue with self-administered Web surveys is how to handle items the user skips. In this study skipped questions are not allowed. When the user clicks Next to continue and an item in the current page was not answered, the page automatically returns with a red symbol and the message "You forgot to answer this question. Please answer it now."

Figure 2: Error message for skipped item
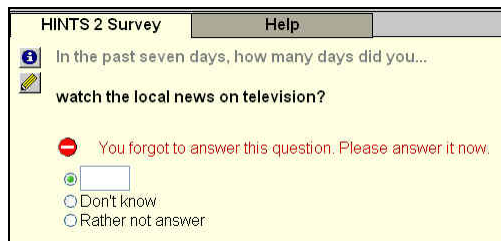
Help and Comments

When the info button [] is pressed, the text explaining the item immediately displays in the area between the question text and the answer choices. If the button is clicked a second time the explainatory text disappears. Similarly, when the comment button []is pressed, a text box appears for entry and the comment is stored as a standard Blaise remark.

Figure 3: Question-level help
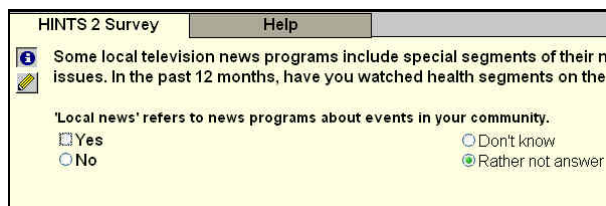
Figure 4: Entry of a comment
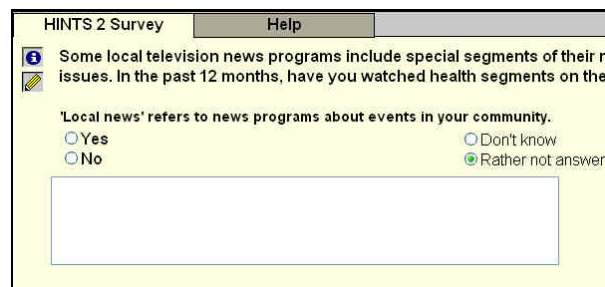
Multiple-item Pages

The system allows displaying multiple questions on the same page. In this study, however, the standard approach was to display one question on each page. Multiple-item pages were implemented only when a follow-up question was needed.

Figure 5: Multiple items on a page
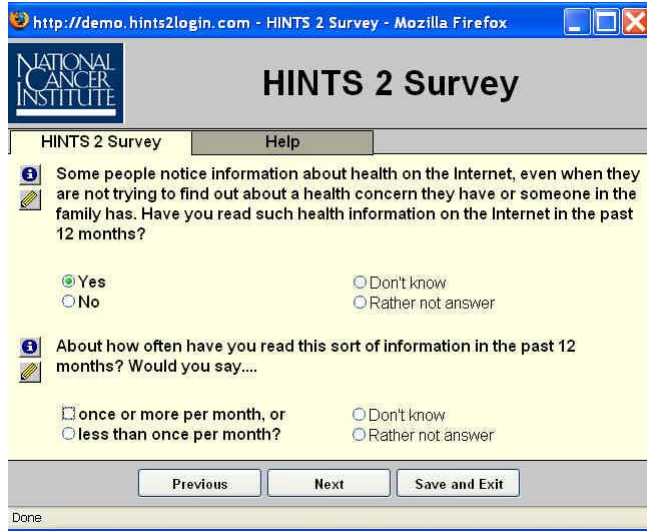
Don't Know and Refusal

The IS system allows Don't Know and Refusal responses to be displayed in a number of different ways to make clear visually that these specialized responses are different from the substantive choices to the question. We decided to display the Don't Know and Rather Not Answer responses in the second column in a less prominent style.

Break-offs and Restarts

In the HINTS 2 application, the 'Save and Exit' button is provided to allow users to exit the application before completion with the hope they will resume it later. If the selection is confirmed, a final receipt page tells that the responses have been saved for the user's return visit.

Figure 6: Dialog on exiting an incomplete interview
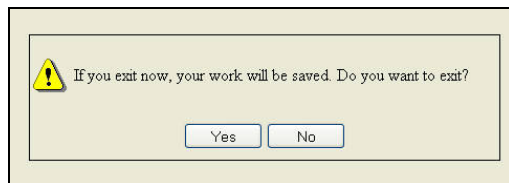
Figure 7: Response page on exiting an incomplete interview

On entering with a valid username/PIN, a partially completed interview is resumed at the point of exit. If one attempts to reenter a completed interview, a message explains the survey has been completed and denies entry.

When a user session is abandoned, the session remains active until the server time-out interview, typically 10 to 20 minutes. When the time-out limit is reached, IS closes the interview session, saving all entries received.

Compound items

Quantity-unit and similar types of compound items are challenging for Web survey respondents. Blaise IS arranges the elements to assist understanding.

Figure 8: Quantity-unit
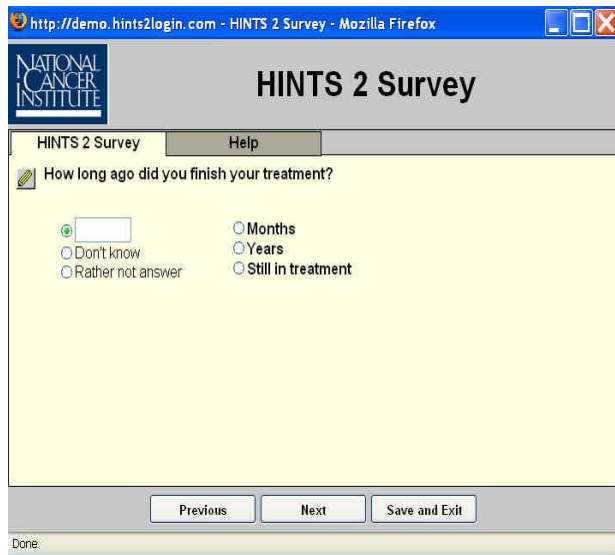
Figure 9: Age or year of occurence
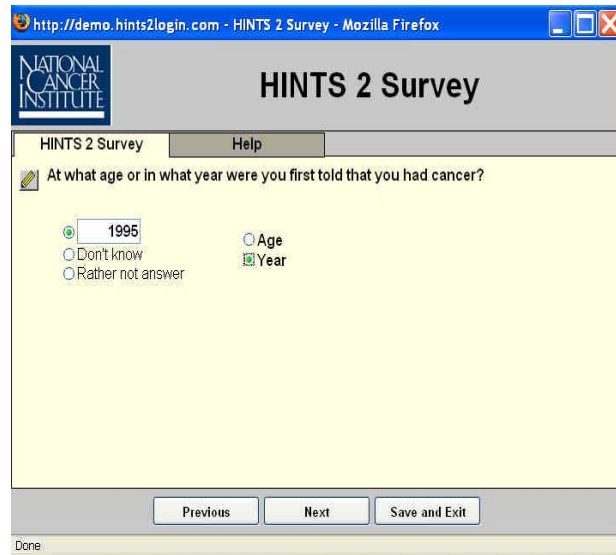
Pop-up Blockers

IS surveys are intended to operate in a pop-up version of the browser in which the normal browser menu elements and buttons are not present. When a pop-up blocker is running on the user's browser, IS detects this and sends a page asking the user to click a button to move forward. As a user-action this opens the custom IS version of the browser. We don't believe this caused any significant difficulties in the HINTS 2 study. But it is somewhat of an inconvenience, albeit a reasonable one given the standardized browser look-and-feel it enables.

Server-side Actions

Blaise IS Web functions are handled by active server page (ASP) processes. The most important ones are:

Startup page	This is the entry point to the survey. In HINTS 2 and most IS surveys, user authentication is the primary function. The user enters the username and password and when submitted to the surver the entry is checked against the case management database.
	If the case is to be interviewed, code in the page takes the appropriate actions. Parameters for the case are generated —starting a new interview or resuming a previous one; passing parameters to guide the interview, for example, the language to start with or parameter values for interview fields.
	Finally, the Blaise IS inteview is activated, driven by the generated parameters.
Receipt page	After the IS interview is completed, the Receipt ASP page runs. Typically, case management-related functions are completed. The interview result and other administrative information are accessed from the interview records. This information is passed to the case management database.
	Finally, a page is sent to the user browser with a thank you and any other information the study wants to provide.
Abort page	The Abort page is run when an IS interview terminates abnormally. This occurs when the session is inactive beyond the allotted period for the server, typically 15 or 20 minutes. The

	Abort page return a message to the user browser explaining that the interview data has been saved and the session terminated. Case management functions are also often implemented here.

Adolescent Trials Network

The Adolescent Medicine HIV/AIDS Research Network (ATN) conducts research related to HIV-infected and HIV-at-risk adolescents, ages 12 through 24 years. It is sponsored by the National Institute of Child Health and Human Development (NICHD).

For the work discussed here The survey is interviewer-administered (CAPI), not self-administered as with most Web surveys. Subjects are interviewed during clinic visits and return at three month intervals to be reinterviewed. Following the initial interview, they are given an immediate feedback report. Blaise IS was chosen because .

- o A simpler model for CAPI, eliminating the logistics of distributing and supporting laptop system and the related management and security issues with distributed interview data

- o Very lengthy survey instrument

- o Able to build on HINTS experience

- o Limited simultaneous use predicted

- o All sites have fast Internet access.

The Blaise IS application has been in the field since May 2005, operating in five clinics. The survey has 509 questions. The volume in the first three months is about 25 interviews per month.

ATN Interviewer Home Page

The ATN web application's startup page is controlled by the ATN web site. It is programmed in ASP following the process described above for HINTS 2. The interviewer enters the interviewer ID, site ID, subject's ID and visit number. Then the IS application is launched.

Figure 10: ATN login page

Web Survey Presentation

The survey user interface is largely the default IS presentation. Special look-and-feel elements are limited since the user is a trained interviewer. Examples are below.

Figure 11: Gender and lead-in question on currently prescribed medications series

Figure 12: Select currently prescribed medications

Figure 13: Dosage for selected medications

Figure 14: Edit for consistency of age and date of birth

ATN Feedback Report

A key requirement of the ATN interview system is the preparation of a real-time report for the subject based on numerous fields from the interview. On completion of the interview the system

- o   extracts data from Blaise,

- o   formats the data as XML,

- o   writes to a SQL*Server table,

- o   reads and formats the reports using a SAS program and

- o   downloads a PDF report for the subject.

The report includes 10 pages summarizing responses on key subjects. Example pages are below.

Figure 15: Problems because of drug and alcohol use

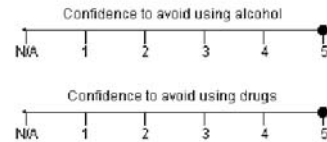**Figure 16: Attitudes about drugs and alcohol**
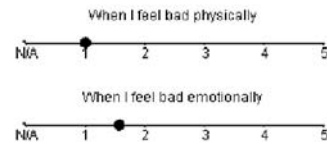
Conclusion

Although these two projects had relatively low numbers of Web survey users, they are substantial implementations in virtually every other respect. The questionnaires are longer and more complex than most. The processes of authentication, management, and interviewing are the same as would be used for larger-scale surveys. The ATN implementation seems particularly cost-effective compared to the logistical issues of distributing and managing disconnected systems in the field sites.

Due to the relatively small number of users, the projects did not run into any problems of scalability (the number of simultaneous Web users the system can support). Scalability will vary based on a number of factors, such as server capability, questionnaire size and complexity, and any special server-side handling requirements. As our experience with the software increases, we expect to refine our understanding of Blaise IS scalability under different models, as well as the effect of various configuration options on performance under high-usage conditions.

We are pleased with our first two experiences with Blaise IS. The start-up burdens were not onerous. The interviewing implementations were very satisfactory. And the overall experience has made us quite eager to work further with Blaise IS in a wide range of Web surveys.