Journal of Official Statistics, Vol. 16, No. 3, 2000, pp. 255–271

Developing Usability Guidelines for AudioCasi Respondents with Limited Literacy Skills

Sid J. Schneider[1] *and Brad Edwards*[1]

In an audio-enhanced computer-assisted self-interview (audioCasi), respondents see questions on a computer screen while they hear them over headphones. Respondents with limited literacy skills potentially could benefit from the ability of audioCasi systems to present aural and graphical cues as well as text. The article describes the development of software for administering audioCasi surveys to both proficient readers and respondents with limited literacy skills. The initial version of the software was designed in accordance with principles drawn from the literature on computer-assisted interviewing and the human-computer interface. The software then underwent seven rounds of usability testing, a process in which people are observed using the software and then debriefed, with the goal of identifying deficiencies in the design of the software. People with a range of literacy skills participated in the usability tests. The software was revised following each round of testing. In the final version of the software, the screens were uncomplicated and the procedures involved in taking a survey were consistent and self-evident.

Key words: Software evaluation; computer-assisted survey information collection.

1. Introduction

The results of the National Adult Literacy Survey suggest that the reading proficiency of about 30 million adult English-speaking Americans is at the lowest defined level. There are various reasons for these individuals' limited literacy skill. For example, approximately one fifth have visual difficulties, and about one quarter have cognitive or health conditions that limit their reading ability (Kirsch, Jungeblut, Jenkins, and Kolstad 1993). Persons who score at this literacy level would have trouble participating in any surveys that required reading, such as paper-and-pencil surveys or computer-assisted self-interviews. Researchers who have needed to survey respondents with literacy problems have usually developed spoken, face-to-face versions of their surveys (e.g., Ferraz, Quaresma, Aquino, Atra, Tugwell, and Goldsmith 1990; Lasater and Mehler 1998).

However, face-to-face interviews can be expensive. Also, they may not be the best method for collecting data on health risk behaviors or other sensitive matters, because respondents are often reluctant to divulge personal information in face-to-face interviews (Tourangeau and Smith 1996, 1998). Persons with literacy problems tend to be in the

[1] Westat Inc., 1650 Research Blvd, Rockville, MD 20850, U.S.A. E-mail: schneis1@westat.com.
Acknowledgments: The authors thank Frank Conner, Ronna Cook, Jim Covell, Mark Gallagher, Beverly Geline, Elizabeth Hintz, Pat Montalvan, and Christopher Roberts. Reprint requests and other correspondence should be directed to Sid J. Schneider.

lower socio-economic strata, a demographic category associated with relatively increased health risk factors (Fournier and Carmichael 1998; Waitzman and Smith 1998). There is a need for a method other than face-to-face interviews to conduct surveys of people who have literacy problems.

A potential solution to this problem has emerged recently with the development of the audio-enhanced computer-assisted self-interview (audioCasi). In an audioCasi, a computer-recorded voice recites each question in a survey while that question appears on a computer screen. A respondent who cannot read well can rely upon the audio component of the survey to present the questions and to explain how to answer. Recent research has suggested that respondents are more willing to report sensitive information in an audioCasi than in a face-to-face interview (Turner, Ku, Rogers, Lindberg, Pleck, and Sonenstein 1998; Yagelka, Zingaro, Spier, Campbell, Vangelatos, and Gjonbalaj 1998) or in a self-administered paper-and-pencil questionnaire (O'Reilly, Hubbard, Lessler, Biemer, and Turner 1994; Turner, Forsyth, O'Reilly, Cooley, Smith, Rogers, and Miller 1998).

AudioCasi is an emerging tool currently moving from experimental studies and field tests to an increasing production role (Miller 1997). Survey developers have been required to design audioCasi applications that can be easily used by people with limited literacy skills, without appearing cumbersome or simplistic to respondents who are proficient readers (Couper 1998). The standards for the design of audioCasi surveys have yet to be defined.

AudioCasi design standards could be developed through a process called ''usability testing.'' The present article describes the initial design and the usability testing of audio-Casi software. The goal of this work was to establish the design standards for audioCasi surveys that can be administered to a general population, in which a substantial proportion of respondents have limited literacy skills.

Usability. In the past several years, many designers of computer-assisted interviews have applied the principles of human-computer interface design to ensure that their software was usable by interviewers and respondents (Bosley, Conrad, and Uglow 1998; Couper 1995; Hansen, Fuchs, and Couper 1997). The term ''usability'' refers to specific, observable software characteristics, including learnability, efficiency, memorability, error handling, and user satisfaction (Hix and Hartson 1993; Neilsen 1993). ''Usability testing'' is a technique for assessing these characteristics in order to identify and remedy the shortcomings of software products.

Learnability refers to the time and effort that a user who is unfamiliar with a software product must invest in order to understand how to use it. In the case of audioCasi, users are likely to expect that the interview procedure will be similar to a face-to-face interview or paper-and-pencil questionnaire, even though the computer-assisted format may be novel. Computer-assisted interviewing systems that meet this expectation are likely to be easy-to-learn, allowing the respondents to concentrate on taking the survey without spending much time learning how to operate the software.

Efficiency refers to the ease with which expert users of a software product can add to their expertise. In the case of audioCasi, efficiency refers to the ability of users to understand all possible situations that may arise. For example, a user might understand how to respond to ''choose one'' multiple choice questions but then encounter a ''choose all that apply'' multiple choice question. The efficiency of the system is a measure of the users' ability to identify the change in the requirements and then to carry out the needed actions.

Memorability refers to the users' ability to recall the procedures for using the software without repeated explanations or instructions. AudioCasi software is likely to be memorable if it maintains a consistent appearance throughout a survey, and user actions elicit predictable behavior from the software. For example, users are likely to remember the action required to answer questions if that action does not vary, regardless of whether the question calls for selecting a multiple choice alternative, or entering a number or a date, or providing some other response. Memorability is a crucial software characteristic for persons with limited literacy skills, because written instructions or help screens, no matter how well designed, can be of little use to respondents who cannot read that material easily.

Error handling refers to the ability of a software product to avoid error situations, and to guide the users to recover from any error situations that do occur. An audioCasi, no less than any other survey technique, is likely to elicit errors such as omitted responses, out-of-range responses to number questions and impossible responses to date questions. When such errors occur, the software must interact with users to explain the nature of the error and the actions required to solve the problem. Error handling is important for users with limited literacy skills because they might blame an error on their lack of reading skill and conclude that they cannot resolve the error by themselves. The software must have a consistent way of resolving errors without requiring the respondents to read anything.

Subjective satisfaction refers simply to how pleasant it is to use the software. Satisfied users feel that the software has helped them, while dissatisfied users feel that the software has imposed unnecessary obstacles. In the case of computer-assisted self-interviews, users are likely to be satisfied if they feel that the interview was as easy-to-take as any face-to-face or paper-and-pencil interview. Many persons with limited literacy skills expect to have difficulty using computers (Couper and Rowe, 1996). Usable software will quickly give these users confidence that they can handle anything that occurs.

Measuring usability. The usability of a software product is assessed with a variety of techniques, including expert reviews, ''think aloud'' interviews, and focus groups (Neilsen 1993). Another usability testing method involves observing and debriefing users actually engaged in learning, using, and interacting with the software (Dumas and Redish 1993; Neilsen 1993; Rubin 1994). Sometimes researchers collect quantitative data about a software product, such as the speed and accuracy with which users carry out specific tasks (Caspar and Couper 1997). These data can help to assess respondent performance, or compare two or more software designs (Neilsen 1999).

However, researchers usually approach usability testing as qualitative, observational research that provides insight into deficiencies in the software design. The software is repeatedly revised and retested in successive rounds, each involving a small number of users. The software may evolve rapidly as usability problems are identified and corrected. This iterative process of testing early and often makes the user the central consideration in the design of the system-user interface.

Software designers begin only with their best guess of what the system-user interface should be. They can base their guess on existing, successful software products, the published research literature, and the opinions of other software designers. However, the best guesses of software designers will not always lead to usable software products. Nor is it necessarily helpful to ask potential users how they would like the software to be designed. Users are unlikely to be able to articulate their preferences until they have had experience using prototypes of the software. Usability testing is based on the idea that improvements in the user interface can be identified by examining the interactions between users and the software (Rubin 1994).

For example, Hansen, Couper, and Fuchs (1998) subjected a computer-assisted personal interview to several evaluation techniques. Usability testing helped to identify the aspects of the system, such as the required use of certain function keys, that caused interviewers to commit errors or fail to ask questions completely and accurately. Keystroke analysis did not uncover all of these problems. The usability testing provided direct evidence about the nature of existing problems, enabling the research staff to reach consensus quickly about needed revisions.

However, usability testing cannot demonstrate that one approach to software design is the ‘‘best.’’ Software products are usually intended for use by people with a very wide range of experiences and preferences. Usability testing is intended to help achieve one of many possible acceptable designs.

2. Method

Usability problems are frequently revealed by small cues, such as a few respondents' momentary hesitations or perplexed looks. For that reason, usability laboratories have equipment that allows the researchers to observe respondents closely.

For this test, the usability laboratory consisted of two rooms separated by a one-way mirror. Respondents participated one at a time. They sat at a table, wearing headphones, in front of a laptop computer. While the respondents used the system, cameras recorded their bodies, faces and hands, a scan converter recorded the computer screen, and a quad splitter combined these inputs into a single split screen image. Table 1 lists the equipment used. Observers watched from behind the one-way mirror, and reviewed the videotapes afterward, noting the respondents' strategies, errors, and moments of vacillation or confusion. All equipment except the ceiling mounted cameras and the one-way mirror could be transported and set up outside the laboratory. For tests at field locations, the ceiling mounted cameras were replaced with a second miniature camera, and no observation room was used.

At the beginning of each session, a researcher described the system and the usability testing to the respondent, and obtained the respondent's signature on an informed consent form. The audioCasi system presented a brief tutorial to explain the system. Then, the audioCasi software presented a survey with sections on demographics, education, housing,

Table 1. Equipment in the usability laboratory

Two Hitachi HV-C 10A CCD cameras on Camrobotic Systems Model 1000PT ceiling mounts
One Provideo KCC-272N miniature camera
One Javelin Pro JQC-456 image splitter
One Sanyo SRT-500 video cassette recorder
One Focus Enhancements TView Gold scan converter

Table 2. Usability testing rounds

Round	Location	Males	Females	Population	Age (if not 18–65)
1	Usability laboratory	1	3	Food pantry users	
2	Shelter for adolescents	4	1	Shelter residents	All <18
3	Usability laboratory	3	2	Food pantry users	
4	Usability laboratory	2	3	General public	One woman 78
5	Usability laboratory	4	2	General public	One woman 81
6	Food pantry	4	0	Food pantry users	
7	Women's shelter	0	5	Shelter residents	

use of medical and dental services, sexual behavior, and alcohol and drug use. The tutorial and survey lasted approximately thirty minutes.

After the respondent finished the audioCasi survey, the researcher conducted a debriefing interview, to identify the features of the audioCasi that were adequate and the features that were confusing or difficult. The interviewer asked the respondents whether they felt confident that they actually entered their intended answers. The interviewer also asked the respondents how they arrived at their procedures for answering the questions, and about alternate designs that might eliminate the difficulties that they encountered.

There were seven rounds of usability tests, involving 34 respondents (Table 2). The adult respondents with limited literacy skills were recruited from a food program for the needy and from a women's shelter. Adolescent respondents with literacy problems were recruited from a runaway shelter. Proficient readers were recruited from the general public and included elderly people and persons with little computer experience. None of the respondents was physically or visually challenged.

It was decided to forego administering literacy tests to the respondents. Such tests could be an initial negative experience for the respondents with limited literacy skills, who might then expect to have further problems with the audioCasi. Instead, the staff of the shelters and the food program identified respondents who had literacy problems.

The software was revised following each round of testing. The seventh round revealed no further salient deficiencies, so no additional rounds were scheduled.

3. Overall Approach to the Software Design

General considerations. The audioCasi software was designed for PC-compatible laptop computers, the hardware usually used to administer audioCasi surveys in the field. The software was developed using Asymetrix Toolbook II, a multimedia application authoring system. The initial design of the audioCasi screens was intended to be consistent, uncomplicated and self-explanatory, so that respondents could learn a minimum number of easy-to-comprehend procedures that would serve them throughout a survey, regardless of the nature or complexity of any question (Bosley, Conrad, and Uglow 1998; Hansen, Fuchs, and Couper 1997). For example, each screen displayed only one question (e.g., Figure 1). There was therefore never any need to explain the number of questions posed on a screen, in keeping with Jenkins and Dillman's (1997) guideline for paper-and-pencil questionnaires (''Ask . . . only one question at a time''), and with Schaefer and Dillman's (1998) finding that item nonresponse is minimized in electronic mail surveys when the number of questions on a screen is minimized.

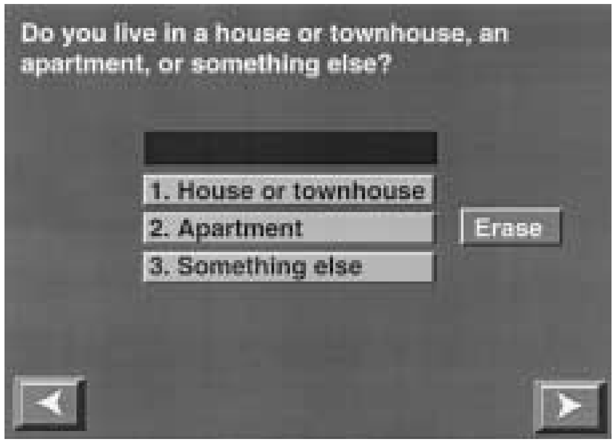

Fig. 1. An example of a ''choose one'' multiple choice question

Nine events, listed in Table 3, could possibly take place in the interaction between a respondent and the audioCasi software. Only four kinds of screen objects were needed to handle all of these events: text, response buttons, response fields and navigation buttons. Text was needed to present the questions, explanations and error messages. Response buttons allowed the respondents to enter a response. Response fields displayed the responses that had been entered. Navigation buttons allowed the respondent to move from screen to screen.

As with any audioCasi, when a screen first appeared, the system recited the question or text item on the screen. The text item or the question was therefore positioned at the top of the screen, the part of the screen that users were likely to see first (Jenkins and Dillman 1997). After the audioCasi presented the question, the respondent answered. Therefore, the response buttons appeared below the text of the question, in the area that the users were likely to see second. The response field appeared just above these response buttons. Thus the response buttons and response fields were grouped together, to suggest they were both relevant to the response to the question (Dillman 1978; Jenkins and Dillman 1997). After answering, the respondent proceeded to the next screen. The navigation buttons were therefore placed at the bottom of the screen, where they would be encountered last.

Color and font. As few colors as possible were placed on any screen. The colors were not used decoratively but rather to help distinguish the various objects on the screen, in

Table 3. Respondent-system interactive states

1. System presents question or text item.
2. Respondent asks to hear item again.
3. Respondent gets more information.
4. Respondent enters response.
5. Respondent changes response.
6. Respondent registers response.
7. System reacts, if response is invalid.
8. Respondent moves to next item.
9. Respondent returns to prior item to check or change response.

keeping with usability standards calling for a functional, minimalist design (Neilson 1993).

At first, there was some doubt about the particular colors that should be used. There has been some research to suggest that the text of the question should appear in black letters on a white background because this combination is very legible and works best for persons with diminished visual acuity (Lynch and Horton 1997). However, there is also research to suggest that the contrast between black letters and a white background might be fatiguing over time (Hix and Hartson 1993). Computer monitor users tend to prefer text to be presented upon a blue or bluish cyan background (Hix and Hartson 1993; Pastoor 1990). Therefore, the screens were designed with white letters on a dark teal background, the combination found on many classroom chalkboards.

Only one font style was used on the screens. This decision allowed the screens to appear as simple as possible, and eliminated the possibility that some users would incorrectly attach some meaning to changes in the font style. Arial, a simple, sans serif font, was chosen. Both capital and lower case letters were used. This typeface is easily read on computer monitors (Lynch 1994). The debriefings following each usability session suggested that these initial color and font choices did not need to be revised.

Touch screen. Respondents interacted with the system by touching the buttons and the text on the screen with a finger. Touching the screen was equivalent to clicking the left button of a mouse while the cursor was upon the area being touched. The respondents were not given a mouse and the keyboard of the laptop computer was covered up.

The decision to use a touch screen was based on the idea that some users, particularly those with limited literacy skills or computer experience, might find a mouse or keyboard to be tedious or foreign. It may be particularly difficult for those respondents to enter dates or to change responses by entering codes on a keyboard. Teaching respondents with little computer experience to touch something on the screen probably involves less explanation than teaching them to locate keys or to use a mouse (Hix and Hartson 1993; Shneiderman 1997). Touch screens are successfully used in information kiosks in shopping malls, museums and airports, suggesting that they are easily learnable for a wide range of respondents.

However, this assumption is challenged by published reports about audioCasi that did require use of the keyboard or a mouse (Turner, Ku, Rogers, Lindberg, Pleck, and Sonenstein 1998). Therefore, the respondents in the present usability tests were observed closely to determine whether they quickly mastered the touch screen technique, and whether they maintained that mastery throughout the session. It appeared that all respondents, regardless of their literacy skills, performed satisfactorily with the touch screen. In particular, the respondents were able to use the touch screens early in the survey, suggesting that the respondents learned the touch procedure quickly. During the debriefings, the respondents who could not read well stated quite unanimously that they would prefer the touch screen over a keyboard or mouse. Respondents who said that they initially feared the computer format reported surprise at the ease with which they mastered the touch screen.

Most commercially available software applications react to mouse button clicks when the button is released, not when the button is first depressed. In these usability tests, however, many respondents stated that they expected the software to react immediately when they touched the screen, not when they lifted their finger. Therefore, the system was re-programmed to react to ''finger down'' not ''finger up.''

Voice. The initial design of the audioCasi software called for a female voice in the audio component of the system. Research with face-to-face interviewers has suggested that respondents of either sex may be more willing to disclose personal information to female interviewers than to male interviewers (Johnson and DeLamater 1976; Pollner 1998), although the literature is not unanimous on this issue (Catania, Binson, Canchola, Pollack, Hauck, and Coates 1996). Research on audioCasi has suggested that the gender of the computer-stored voice has almost no impact on respondent behavior, even in surveys on sensitive topics (Turner, Forsyth, O'Reilly, Cooley, Smith, Rogers, and Miller 1998).

The speaker attempted to sound nonjudgmental and pleasant, as though she was speaking directly to the respondent, rather than announcing to a large audience. She also attempted to avoid any accent or excessive inflection.

In the usability tests, the adolescent respondents had the strongest opinions about the voice. They objected to one speaker whose voice they found to be lilting and variable in tone. The respondents in the women's shelter stated that they would prefer a female voice over a male voice. The other respondents stated that they had no objection to the female voice, even for the sensitive questions. As a result of this feedback, a particular female speaker was selected whose voice was acceptable to all of the respondents who heard it.

Navigation buttons. Only two navigation buttons were needed: one to proceed to the succeeding screen (called a ''next'' button), and one to return to an earlier screen (called a ''back'' button). The navigation buttons were designed with no text on or near them. They could be readily recognized by their distinctive location, shape, size and color. They had a light blue color with a speckled texture. On every page, the ''next'' button appeared at the bottom right corner and had an arrow pointing right. The ''back'' button appeared symmetrically on the bottom left corner and had an arrow pointing left.

When the respondent pressed the ''next'' button, the audioCasi software evaluated the respondent's answer. If the answer was unacceptable, the software presented an error message, described below. Otherwise, the audioCasi presented the next screen in the survey. There was a transition effect, provided by the Asymetrix Toolback system, that created the illusion of a page turning from the right to the left. This metaphor of turning pages suggested that the survey was proceeding forward.

When the respondent pressed the ''back'' button, the audioCasi presented the screen that the respondent saw immediately before the current screen. The transition to that screen created the illusion of a page turning from the left to the right, returning to an earlier page. When the earlier screen appeared, the respondent's answer to that item was still visible.

The respondents did not appear to have any difficulty with the navigation buttons during the usability tests. The debriefings suggested that all respondents, including those with literacy problems, were able to discern the function of the buttons without resorting to trial and error. For that reason, the design of the buttons was not revised during the testing.

Error messages. The error messages were in bright green letters, appearing in the lower center of the screen between the ''next'' and ''back'' buttons. The audioCasi voice recited the error message when it appeared. The error message vanished the next time the respondent touched the screen, such as to change the response.

The usability tests suggested that the error messages written for the initial version of the system created confusion. Those messages described the nature of the error but failed to present the actions that the respondent had to perform to resolve the error. Therefore a large number of error messages were written so that the system could describe as precisely as possible both the error condition and the action necessary to resolve the error. The error messages avoided assigning blame to the respondent with words like ''You failed to . . . '' Rather, the messages guided the respondent to an appropriate corrective action, in accord with Shneiderman's (1997) guidelines.

Responses to Questions. The response buttons were distinguishable by their appearance. They were all rectangular, light gray, with a smooth texture and black lettering. Their shadowing made them appear to protrude slightly, like telephone buttons. When respondents pressed response buttons, the audioCasi displayed the response in the response field in white letters on a black background.

The audioCasi presented only visual, not audio feedback when the respondents selected a response. For example, the system never recited ''You have chosen the 'no' response.'' It was thought that audio feedback would introduce unneeded complexity, especially for proficient readers. The usability test confirmed that respondents did not desire audio feedback. The respondents were much more concerned that they could understand the content of the response buttons, so that they could feel confident that they were entering their intended response.

Controlling the voice. Respondents could hear a question again by touching the text of the question. Also, they could interrupt the voice by touching anywhere on the screen. For example, if the audioCasi was reciting an error message and the respondent touched the text of the question, the audioCasi interrupted its recitation of the error message and recited the question again. Or, if the audioCasi was reciting the question and the respondent touched a response button, the audioCasi interrupted its recitation of the question and fell silent. In this way, proficient readers could respond to a question without having to wait for the audioCasi to finish reciting it. These features gave the respondent control over the voice, and provided a shortcut through a potentially long procedure, in keeping with Shneiderman's (1997) guidelines.

4. Design of Specific Question Types

Each screen in the audioCasi was designed to conform with the overall approach to the software design outlined above.

Text Screens. An audioCasi could contain text screens to introduce topics and questions and to provide explanations. Text screens had no response fields or response buttons. They had only text and navigation buttons.

''Choose one'' multiple choice questions. In ''choose one'' multiple choice questions (Figure 1), the response buttons were arranged vertically, below the response field. Because respondents with literacy problems could not be expected to read the response buttons, the audioCasi was programmed to recite the response alternatives. When a screen with a multiple choice question appeared, the audioCasi first recited the question. It then recited the first alternative, while the background on the response button showing the first alternative turned yellow. Then, that button returned to gray while the response button having the second alternative turned yellow and the audioCasi recited the second alternative. The audioCasi continued in this fashion until all the alternatives were presented. If the respondent touched the question, the audioCasi recited the question again, and then recited and highlighted each alternative successively again.

The usability tests suggested that respondents with limited literacy skills used a number of cues to enter their response. They listened to the voice reciting the alternatives and watched the alternatives being highlighted. They noted the numbering of the alternatives, when it was present, to differentiate the choices. However, the respondents did not tend to compare the captions on the response buttons with the contents of the response field to ensure that they made the choice that they intended. The usability tests confirmed that the cues that these respondents did use enabled them to make their intended choices.

Number Questions. On screens having questions calling for a numeric response, the response buttons were arranged like the numbers 0 through 9 on a telephone keypad. That is, the lower numbers except zero appeared on top, not at the bottom as they do on a calculator keypad. The response field appeared just above this keypad. A large proportion of persons with limited literacy skills can identify numbers (Kirsch, Jungeblut, Jenkins, and Kolstad, 1993). For that reason, the system was not programmed to recite the individual numbers or to highlight the response buttons.

If a respondent filled the response field with digits and then pressed an additional button in the number keypad, the leftmost digit in the response field vanished, the rest of the response shifted to the left, and the newly chosen digit appeared in the rightmost position. In this way, pressing a number response button always brought about the same action – a digit was added to the number in the rightmost position in the response field, regardless of whether or not the response field was full.

Usability testing confirmed that respondents who could not read well could read numbers and enter their responses successfully.

Quantity and unit questions. Quantity/unit questions required both a number and a multiple choice response. For example, the question ''How often do you drink a glass of beer?'' could receive a response of ''8 times per day'' or ''10 times per week'' or ''3 times per month.'' The screens for these questions had a number keypad and a response field to the left, and a set of multiple choice response buttons and another response field to the right (Figure 2). When the audioCasi presented a quantity/unit question, it recited the text of the question. Then, the multiple choice response buttons were highlighted in yellow successively while the respective texts on the buttons were recited. In this way, the audio-Casi presented the alternative responses in the manner that it presented the alternatives in multiple choice questions.

After a few rounds of usability tests, it was apparent that this kind of question was confusing to some respondents. However, the tests did not clearly suggest a more promising design for such a question. Little improvement was noted when changes were made in the instructions given when the question was recited. Therefore, an explanation for these questions was added to the tutorial at the start of the audioCasi and again immediately before the first quantity/unit question that appeared in a survey. Also, the error messages associated with this kind of question were expanded, with more

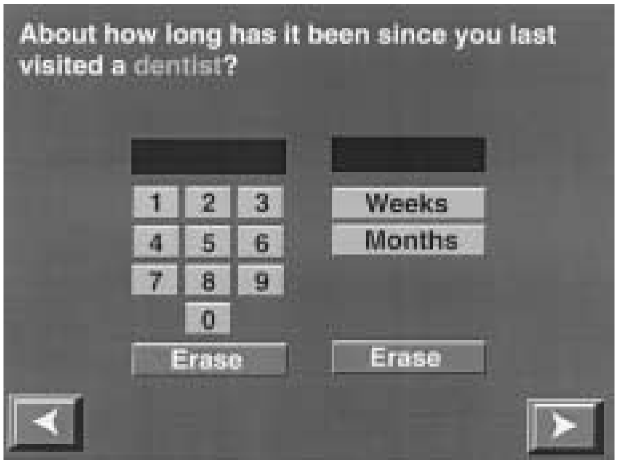

Fig. 2. An example of a quantity/unit question

explanation of how to resolve omissions of the quantity or the unit component of the response. These changes helped respondents to learn this kind of question and to resolve errors.

Date questions. All date questions required a month keypad, consisting of response buttons arranged in a 3 by 4 layout. The months were presented as three-letter abbreviations. Only the first letter of the abbreviations was capitalized. The earliest months of the year appeared at the top of the keypad.

Month-only questions had one keypad and one response field. Month and day questions and month and year questions had two keypads and two response fields side by side. Month, day and year questions had three keypads and three response fields side by side. When a respondent pressed a month response button, the full name of the month appeared in the response field. When the respondent pressed a day response button or a year response button, the number on the button appeared rightmost in the response field. A maximum of two digits could appear in the response field for the day.

Usability testing showed that the respondents with limited literacy skills could recognize the abbreviations for the months of the year. They were also familiar with the month-day-year format for dates. Therefore, there was no need for the audioCasi software to recite the names of the months or explain the date format.

''Choose all that apply'' questions. ''Choose all that apply'' multiple choice questions were presented much like ''choose one'' multiple choice questions: first the system recited the question, and then it highlighted and recited each successive response alternative. To make ''choose all that apply'' questions strictly consistent with the other questions, all selected response alternatives would have to appear in the response field, located above the response buttons. However, the response field was too small to display more than a few selected responses.

It was therefore decided to make the appearance and behavior of this question inconsistent with those of the other kinds of questions. The response field was eliminated in ''choose all that apply'' questions. Instead, there was a check box to the far left on each response button. When the respondent pressed a response button, a check appeared in the check box on that button. If the respondent pressed that button again, the check vanished.

Each ''choose all that apply'' question had a ''none of these'' response button, located below and separate from the other response buttons. When the respondent pressed the ''none of these'' button, a check appeared in the check box on that button, while any checks in the check boxes on any other response buttons on the page vanished. When a respondent pressed any response button other than the ''none of these'' button, any check in the check box of the ''none of these'' button vanished.

The absence of a response field and the presence of the check boxes was intended to help respondents distinguish between ''choose one'' and ''choose all that apply'' multiple choice questions. Usability tests suggested that these visual cues diminished, but did not eliminate, that confusion. It was helpful to put an explanation about ''choose all that apply'' questions in the tutorial and again immediately before the first such question (Dillman 1978). It was also helpful to emphasize in the text of every ''choose all that apply'' question that more than one response was permissible.

5. Other Issues Addressed by Usability Testing

Erase button. All questions required at least one ''erase'' button that respondents pressed to erase any answers that they had already entered. The ''erase'' button would accommodate respondents who selected a response but then decided that they would rather leave the question unanswered. Also, in number questions and ''choose all that apply'' multiple choice questions, the simplest way to change an answer was to press the erase button and then enter the new response. Just as importantly, the ''erase'' button was needed so that respondents would always feel assured that any entry they made could be easily undone (Shneiderman 1997). In usability tests of CASIC software (House and Nicholls 1988), users wanted this capability so that they could quickly eliminate unintended entries and begin anew.

There was one erase button for every response field. Erase buttons appeared just below response buttons. They resembled response buttons, except that they were dark gray rather than light gray, and had white lettering rather than black.

In the usability tests, respondents with limited literacy skills reported that they found the word ''Erase'' to be easier to read than equivalent words like ''Clear.'' However, many respondents reported that they had to remember the function of the ''erase'' button from screen to screen; the function of the button was not self-evident. No icon could be identified to substitute for the word ''erase'' on the button. For example, a picture of a pencil eraser could not convey the meaning of the word ''erase.'' Therefore, the tutorial was expanded to teach respondents the purpose of the ''erase'' button, and to help them rehearse using it.

''Next'' button timeout. In the usability tests, some respondents forgot to press the ''next'' button. They entered a response or listened to the contents of a text page, and then simply waited for something to happen. For that reason, a reminder to press the ''next'' button was added. The reminder appeared when 15 seconds elapsed after a respondent entered a response into every response field on a page or pressed at least one button in a ''choose all that apply'' question, or heard all of the text of a text screen.

The reminder was a special error message (''Please press the button to move on when you are finished'') accompanied for 15 seconds by a flashing green rectangle surrounding the ''next'' button. This reminder was sufficient to resolve this issue.

Tutorials. Usability testing demonstrated that the tutorials were valued by the respondents. The tutorials began with a general description of the audioCasi method, the headphones and the touch screen. The tutorials also explained the concepts relevant to each kind of question used in a survey. They covered the purpose of the text, response field, response buttons, and navigation buttons. The tutorials had the respondents practice answering each type of question.

Additional features. In the final rounds of usability tests, two features were added not because there were usability problems to be fixed, but rather because future survey designers might need them. First, a ''hotword'' capability was added in which the respondents could touch certain words on the screen, identified by their yellow color, and receive an explanation of the words. This explanation appeared between the navigation buttons while it was recited.

Also, a popup screen was designed for questions in which it was necessary to know why a respondent failed to answer. The popup screen contained a multiple choice question: ''You did not answer the question. Why not?'' with three response buttons: ''Don't know the answer,'' ''Don't want to answer,'' and ''Really meant to answer.''

6. Conclusion

The usability tests suggested that respondents with limited literacy skills can complete an audioCasi survey, if the survey software: (1) limits the number of objects on the screen, and the principles governing the actions of those objects, to the absolute minimum possible, and (2) has as much consistency as possible from screen to screen. These standards largely eliminate the need for the respondents to read any text to understand the questions,

Table 4. Characteristics of the initial design of the audioCasi and their refinement through usability testing

Characteristic	Principal Source	Results of usability testing
Minimal number of object types	Neilsen, 1993	Objects modified
Consistent procedures	Neilsen, 1993	Procedures modified
One question per screen	Jenkins & Dillman, 1997	Unchanged
Position and appearance of objects	Dillman, 1978; Neilsen, 1993	Objects modified
Color and font	Lynch & Horton, 1997; Hix & Hartson, 1993	Unchanged
Touch screen	Shneiderman, 1997	Unchanged
Female voice	Turner et al., 1998	Specific voice chosen
Turning page metaphor	Asymetrix Toolbook II	Unchanged
Error messages	Couper, 1998; Shneiderman, 1997	Content revised
Interrupt the voice	Shneiderman, 1997	Unchanged
''Erase'' button	Couper, 1998; Shneiderman, 1997	Appearance changed
Hotwords	Sperry et al., 1998	Unchanged
Tutorial	Dillman, 1978	Content expanded

Table 5. Examples of observations during usability testing and modifications made to the software as a result

Observation	Modification to software
Respondents did not always enter dates in the order month, day, year, creating confusion	Separate response fields provided for month, day, and year
Unit/quantity questions caused respondents to hesitate; some failed to provide both unit and quantity	Separate response fields created for unit and quantity; error messages expanded; more instructions provided before question
Respondents failed to press ''next'' button on text screen	Timed reminder built into screen
Respondents hesitated, made false starts when answering multiple choice question	Multiple choice alternatives presented as push buttons rather than radio buttons
Respondents failed to discern difference between ''choose one'' and ''choose all that apply'' questions	Appearance of ''choose all that apply'' questions altered to suggest that more than one response was acceptable
Respondents still failed to understand ''choose all that apply'' questions	Tutorial enhanced; also, instructions provided with question
Error messages in popup window caused respondents to hesitate when presented	Error messages incorporated into screen, made to vanish automatically
One respondent futilely attempted to erase an entry by repeatedly re-entering the screen	Erase buttons placed on all screens in a consistent way
Respondents confused by the erase button	Instructions on erase button added to tutorial

responses, and procedures in the survey. Table 4 shows some of the components of the initial design of the audioCasi. Table 5 shows some of the usability problems that were observed as the software evolved through repeated testing, and the changes that were made to resolve these problems.

There will always be some features of an audioCasi system that are difficult to explain, even when these standards are followed. This usability test suggested that a tutorial at the start of a survey, and instructions embedded within the survey, could help familiarize respondents with these features. Respondents who cannot read well are amenable to receiving explanations from a computer, in the same way that they might accept explanations from a helpful person who can read proficiently.

The technique used to develop the audioCasi system described in this article, usability testing, is frequently used to develop software, and is increasingly being used in survey design (Couper 1995). Usability testing usually involves collecting observational data from a purposive sample, rather than quantitative data from a probability sample. The goal of usability testing is to arrive at one of the many possible acceptable designs; it is not to identify a single ''best'' design or to draw conclusions that can be generalized to large populations of users.

AudioCasi appears to be a suitable modality for administering surveys to low-literacy respondents, particularly when the surveys pertain to personal or sensitive material. Further testing could suggest whether audioCasi is more appropriate for persons whose lack of literacy skill stems from cognitive difficulties and lack of sufficient education, rather than physical or visual problems. As researchers gain more experience with audioCasi, the standards for their design will become increasingly well defined.

7. References

Bosley, J., Conrad, F.G., and Uglow, D. (1998). Pen CASIC: Design and Usability. In Computer Assisted Survey Information Collection, eds. M.P. Couper, R.P. Baker, J. Bethlehem, C.Z. Clark, J. Martin, W.L. Nicholls, and J.M. O'Reilly. New York: Wiley, 521–542.

Caspar, R.A. and Couper, M.P. (1997). Using Keystroke Files to Assess Respondent Difficulties with an AudioCasi Instrument. Proceedings of the American Statistical Association, Section on Survey Research Methods, 239–244.

Catania, J.A., Binson, D., Canchola, J., Pollack, L.M., Hauck, W., and Coates, T.J. (1996). Interviewer and Question Effects on Sex Items. Public Opinion Quarterly, 60, 345–375.

Couper, M.P. (1995). Discussion: What Can CAI Learn from HCI? Statistical Policy Working Paper 23, Seminar on New Directions in Statistical Methodology. Federal Committee on Statistical Methodology, Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget.

Couper, M.P. (1998). The Application of Cognitive Science to Computer Assisted Interviewing. In Cognition and Survey Research, eds. M.G. Sirken, D.J. Hermann, S. Schechter, N. Schwarz, J.M. Tanur, and R. Tourangeau. New York: Wiley, 277–300.

Couper, M.P. and Rowe, B. (1996). Evaluation of a Computer-Assisted Self Interview (CASI) Component in a CAPI survey. Public Opinion Quarterly, 60, 89–105.

Dillman, D. (1978). Mail and Telephone Surveys: The Total Design Method. New York: Wiley.

Dumas, J.S. and Redish, J. (1993). A Practical Guide for Usability Testing. Norwood, NJ: Ablex.

Ferraz, M.B., Quaresma, M.R., Aquino, L.R., Atra, E., Tugwell, P., and Goldsmith, C.H. (1990). Reliability of Pain-Scales in the Assessment of Literate and Illiterate Patients with Rheumatoid Arthritis. Journal of Rheumatology, 17, 1022–1024.

Fournier, A.M. and Carmichael, C. (1998). Socioeconomic Influences on the Transmission of Human Immunodeficiency Virus Infection: The Hidden Risk. Archives of Family Medicine, 7, 214–217.

Hansen, S.E., Couper, M.P., and Fuchs, M. (1998). Usability Evaluation of the NHIS CAPI Instrument. Proceedings of the American Statistical Association, Section on Survey Research Methods, 928–933.

Hansen, S.E., Fuchs, M., and Couper, M.P. (1997). CAI Instrument Usability Testing. Proceedings of the American Statistical Association, Section on Survey Research Methods, 1023–1028.

Hix, D. and Hartson, H.R. (1993). Developing User Interfaces: Ensuring Usability Through Product and Process. New York: Wiley.

House, C.C. and Nicholls, W.L. (1988). Questionnaire Design for CATI: Design Objectives and Methods. In Telephone Survey Methodology, eds. R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls, and J. Waksberg. New York: Wiley, 421–436.

Jenkins, C.R. and Dillman, D.A. (1997). Towards a Theory of Self-Administered Questionnaire Design. In Survey Measurement and Process Quality, eds. L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin. New York: Wiley, 165–199.

Johnson, W. and DeLamater, J. (1976). Response Effects in Sex Surveys. Public Opinion Quarterly, 40, 165–181.

Kirsch, I., Jungeblut, A., Jenkins, L., and Kolstad, S. (1993). Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey. Washington DC: U.S. Department of Education, National Center for Educational Statistics.

Lasater, L. and Mehler, P.S. (1998). The Illiterate Patient: Screening and Management. Hospital Practice, 33, 163–165.

Lynch, P.J. (1994). Visual Design for the User Interface. Journal of Biocommunications, 21, 6–15.

Lynch, P.J. and Horton, S. (1997). Yale C/AIM Web Style Guide. New Haven: Yale Center for Advanced Instructional Media.

Miller, P.V. (1997). Is ''Up'' Right? The National Household Survey on Drug Abuse. Public Opinion Quarterly, 61, 627–641.

Neilsen, J. (1993). Usability Engineering. Cambridge, MA: Academic.

Neilsen, J. (1999, August 25). Usability Metrics – How Good Are You? ZDNet, www.zdnet.com, accessed September, 1999.

O'Reilly, J., Hubbard, R., Lessler, J., Biemer, P., and Turner, C. (1994). Audio and Video Computer Assisted Self-Interviewing: Preliminary Tests of New Technologies for Data Collection. Journal of Official Statistics, 10, 197–214.

Pastoor, S. (1990). Legibility and Subjective Preference for Color Combinations in Text. Human Factors, 32, 157–171.

Pollner, M. (1998). The Effects of Interviewer Gender in Mental Health Interviews. Journal of Nervous and Mental Disease, 186, 369–373.

Rubin, J. (1994). Handbook of Usability Testing. New York: John Wiley and Sons.

Schaefer, D.R. and Dillman, D.A. (1998). Development of a Standard E-mail Methodology. Public Opinion Quarterly, 62, 378–397.

Shneiderman, B. (1997). Designing the User Interface: Strategies for Effective Human-Computer Interaction. Reading, MA: Addison Wesley.

Sperry, S., Edwards, B., Dulaney, R., and Potter, D.E.B. (1998). Evaluating Interviewer Use of CAPI Navigation Features. In Computer Assisted Survey Information Collection, eds. M.P. Couper, R.P. Baker, J. Bethlehem, C.Z. Clark, J. Martin, W.L. Nicholls, and J.M. O'Reilly. New York: Wiley, 351–366.

Tourangeau, R. and Smith, T.W. (1996). Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context. Public Opinion Quarterly, 60, 275–304.

Tourangeau, R. and Smith, T.W. (1998). Collecting Sensitive Information with Different Modes of Data Collection. In Computer Assisted Survey Information Collection, eds. M.P. Couper, R.P. Baker, J. Bethlehem, C.Z. Clark, J. Martin, W.L. Nicholls, and J.M. O'Reilly. New York: Wiley, 431–454.

Turner, C.F., Forsyth, B.H., O'Reilly, J.M., Cooley, P.C., Smith, T.K., Rogers, S.M., and Miller, H.G. (1998). Automated Self-Interviewing and the Survey Measurement of

Sensitive Behaviors. In Computer Assisted Survey Information Collection, eds. M.P. Couper, R.P. Baker, J. Bethlehem, C.Z. Clark, J. Martin, W.L. Nicholls, and J.M. O'Reilly. New York: Wiley, 455–474.

Turner, C.F., Ku, L., Rogers, S.M., Lindberg, L.D., Pleck, J.H., and Sonenstein, F.L. (1998). Adolescent Sexual Behavior, Drug Use, and Violence: Increased Reporting With Computer Survey Technology. Science, 280, 867–873.

Waitzman, N.J. and Smith, K.R. (1998). Phantom of the Area: Poverty-Area Residence and Mortality in the United States. American Journal of Public Health, 88, 973–976.

Yagelka, J.R., Zingaro, M., Spier, E., Campbell, J., Vangelos, L., and Gjonbalaj, S. (1998). Development and Evaluation of an Enhanced Computer-Assisted Self-Interview (E-CASI) Intake System. Unpublished data, National Development and Research Institutes, New York.

Received October 1998
Revised October 1999