Advances in Nursing Science
Vol. 27, No. 2, pp. 129–137

Computer-assisted Survey Interviewing of School-age Children

Lynn Rew, EdD, RN, FAAN; Sharon D. Horner, PhD, RN, CNS; Lou Riesch, BA, MPH; Roger Cauvin, BS

Longitudinal surveys of children in school settings can yield large volumes of data for use in exploring research questions that may influence development of intervention programs. However, traditional methods of data collection such as paper-and-pencil questionnaires or person-to-person interviewing present major threats to validity throughout the process of data collection, data management, and data analysis. The use of computer-assisted survey interviewing is an alternative that has both advantages and disadvantages. Lessons learned in the first wave of a longitudinal study of approximately 1161 children in fourth, fifth, and sixth grades illustrate both the strengths and limitations of this method. **Key words:** *computer-assisted survey interviewing*, *data management*, *longitudinal survey*, *school-age children*, *validity*

IMPROVING the health status and health behaviors of children and adolescents is a priority set forth in *Healthy People 2010.*[1] Meeting this goal depends in part on developing health interventions that are based on the analysis of valid data from the designated recipients of these efforts. If one wishes to improve the health behaviors of children then one must study children's health behaviors and identify health risks that are amenable to change. Many behaviors that have their origins in late childhood and early adolescence, such as dietary habits, low physical activity, adolescent pregnancy, and drug abuse, can present significant threats to the health status of the nation's future adults.[2] Much that we have learned about health-risk behaviors comes from cross- sectional surveys of youth, such as the annual Youth Risk Behavior Surveillance, which catalogues the health-risk and health-promoting behaviors of adolescents in schools across the United States.[3] However, little is known about how health behaviors develop and change from childhood to adolescence.[4] To learn how health patterns change over time, longitudinal studies are needed.

Longitudinal studies yield rich dividends in data and are the preferred research designs when the focus is on developmental or other changes over time.[5] However, longitudinal designs can be quite challenging to conduct and have inherent limitations. In particular, longitudinal studies are costly in terms of time commitment, human resources, supplies, and other resources. Other limitations of large-scale longitudinal studies are related to validity of data that can result from errors in data entry. Some of these problems are directly related to traditional methods that rely on self-report survey data obtained by paper-and-pencil questionnaires or person-to-person

From the School of Nursing, The University of Texas at Austin (Drs Rew and Horner and Mr Riesch); and Roger Cauvin Inc (Mr Cauvin), Austin, Tex.

This article was made possible by a grant to the first author (R01-5HD39554–01A2) cofunded by the National Institute of Child Health and Human Development and the National Institute of Nursing Research/National Institutes of Health. The authors thank Dann Coakwell for his technical consultation and expertise in the preparation of this manuscript.

Corresponding author: Lynn Rew, EdD, RN, FAAN, School of Nursing, The University of Texas at Austin, 1700 Red River, Austin, TX 78701 (e-mail: ellerew@mail.utexas.edu).

interviews. Computer-assisted survey-interviewing (CASI) has recently been developed to assist investigators avoid the limitations of paper-and-pencil surveys and direct person-to-person interviews and still collect and record valid research data.[6]

With CASI, the research participant reads survey items on a computer screen and indicates his or her response. The responses can be directly stored in a data management and statistical analysis program, thereby eliminating a point for potential human error that occurs with manual data entry. The CASI method has been shown to be superior in a number of ways to paper-and-pencil questionnaires and person-to-person interviews in studies of adults and adolescents.[7-11] However, very little has been published about the use of CASI with school-age children. Therefore, the purpose of this article is to describe a CASI method for data collection, to identify its strengths and limitations, and to discuss how this technology was used in the first wave of data collection for a large longitudinal study of a multiethnic sample of school-age children.

CASI METHODOLOGY

CASI uses interactive technology to replace paper-and-pencil questionnaires and person-to-person interviewing protocols in research.[12] A computer program is written to present individual survey or questionnaire items on the computer monitor screen for the participant to read. The participant uses a person-to-computer interface such as the mouse, touch, or light pen to select the appropriate response. In some cases, audio technology is added so that the respondent can listen to the question through headphones while reading the question on the computer screen.[13] This is especially beneficial for persons with low literacy and those who are easily distracted or bored with the paper-and-pencil format.[6] Additional software programs can be written by a computer program specialist that will allow the participant's responses to be transmitted to a secure Web site with programming for direct data entry into a statistical analysis program such as the Statistical Package for the Social Sciences (SPSS) for data analysis.[14]

Advantages

There are many advantages to using CASI for data collection. These include cost savings, increased privacy of responses, and greater freedom to report behaviors that are deemed to be risky or socially undesirable. These advantages result in more valid and reliable data.

Cost-effectiveness

While the initial cost of purchasing laptop computers, getting Internet connections, and hiring technicians to translate the surveys or questionnaires into computer programs is high, the long-term costs associated with a large research staff needed for traditional data collection and data entry are greatly reduced. The net effect is that CASI is cost-effective in terms of reduced personnel costs, printing costs, reduced loss of data owing to skipping items or responding outside the format range,[15] and greater ease in cleaning data for analysis. When audio technology is added to enhance CASI, this facilitates participation in the study by persons with lower reading skills and those for whom English is not the primary language.

Privacy and anonymity

Previous research with adolescent and adult populations has shown that self-administered questionnaires afford greater privacy and anonymity than do those administered by a person.[15] While the paper-and-pencil method of response affords greater privacy than do a person-to-person interview, some respondents express concern that they can be identified by their handwriting (especially when open-ended questions are included), thus jeopardizing the confidentiality of their responses. The

CASI method assures privacy of responses. Once the study participant selects his or her answer response, the survey item disappears off the monitor screen. The participant avoids handing his or her survey directly back to the research staff as would happen in paper-and-pencil administration. This may provide a level of comfort, especially in the case of school-age children, if they fear being judged by an adult on the basis of their responses. With this increased level of comfort, it is more likely that the child will answer the survey items accurately and honestly.

Increase in reporting of health-risk behaviors

While people may be reluctant to report health-risk behaviors such as smoking, drinking, use of illicit drugs, and promiscuous and unprotected sexual activity in a face-to-face interview or on a survey form where they believe they can be connected to the answers (either through handwriting recognition or because they give the completed survey to a staff member), CASI has been found to increase reports of such behaviors in adolescents.[10,16] For example, Turner and colleagues found increased response rates among adolescents ($n = 1672$) concerning same-sex sexual activity, use of injection drugs, and sexual activity with persons using intravenous drugs when data were collected with CASI versus a traditional self-administered questionnaire. Similarly, in a study of drug use, sexually transmitted diseases, and unplanned pregnancy among vocational students (aged 15–21 years) in Thailand ($n = 1725$), van Griensven and colleagues found that audio-CASI resulted in high enrollment in the study (99% of eligible students), presumably because of greater acceptability of the method of data collection.

Adolescents who participated in a study of substance use and psychological well-being were randomly assigned to either a CASI-only condition or a combined CASI plus a paper-and-pencil self-administered questionnaire.[15] Reports of health-risk behaviors such as smoking, drinking alcohol, using marijuana, and using illicit drugs were significantly higher in the group using the CASI-only method. There was also a significant difference between the groups on a measure of depressive symptoms, with those responding via the computer reporting a greater number of symptoms than did those in the combined-methods group. The researchers also asked a series of questions about how the participants felt about the method of data collection used. Those in the CASI-only group were significantly more likely than those in the combined-methods group to (*a*) approve of the use of a computer for data collection, (*b*) feel assured that responses would be treated confidentially, and (*c*) feel certain that during the interview nobody could tell how they answered sensitive questions about drug and alcohol use.

Several researchers have tested the hypothesis that research participants are more willing to self-report stigmatized or socially undesirable behaviors via CASI when compared to alternative methods. Studies in the adult population found that CASI significantly increased reporting of stigmatized or sensitive behaviors such as drug use and sexual behavior.[12,17] A study that included qualitative data found that the study participants endorsed CASI as a method that was more likely to elicit honest responses.[6] Data collected from over 6000 adolescents aged 15–21 found higher reporting of sensitive behaviors ranging from alcohol use, sexual history, and health history, when using the CASI method.[13]

Accuracy of data entry

The CASI method can eliminate human errors associated with data entry from paper-and-pencil questionnaires or person-to-person interviews.[6] A corollary to this is that the respondent selects only one answer response, and so errors associated with marking 2 or more responses on a form are eliminated. For example, if the participant is directed to provide only one answer and he or she tries to answer twice for the same question the computer will not allow this.

Consequently, there is less room for errors and missing data due to participant error. With CASI, participants' responses can be exported directly to a secure Web site where statistical analysis can be performed, thus eliminating the need for data entry by staff. This direct input of responses results in more valid responses because human error in the data-entry phase is eradicated.[5]

Avoidance of stigma associated with limited literacy

Persons with limited literacy are unable to complete traditional paper-and-pencil surveys unless the questions are read to them. Limited literacy and the associated stigma are certainly potential problems when collecting data from school-age children because of the variable reading skills within different age groups and the impact of learning disabilities on reading proficiency.[18] Reading items not only takes additional staff and data collection time, but also introduces the possibility that the child will select socially desirable responses. By adding an audio component to the CASI, children with limited literacy, including those for whom English is a second language, can complete the survey or interview by using headphones and following along with the written word on the computer monitor.[19] In this way a potentially embarrassing situation is avoided and these children can experience success in the research process on par with their peers.

Reliability and validity

The use of audio-CASI with adolescent populations shows that respondents increase the reliability and validity of data collected in this way. For example, in Turner's and colleagues' survey of high-risk behaviors among 1672 adolescents, they randomly assigned half the adolescents to a traditional paper-and-pencil self-administered questionnaire group and the other half to an audio-CASI technology group.[10] These researchers found significantly higher reporting of engagement in the health-risk behaviors of intravenous drug use, unprotected sexual activity, and male-to-male sexual activities among the respondents who used the audio-CASI method than among those completing the paper-and-pencil questionnaires. Respondents who used the audio-CASI method were also less likely than other respondents to select optional responses of "I don't know" or "I refuse to answer." Moreover, those using the audio-CASI method had fewer missing data (0.5%) than did those completing the paper-and-pencil questionnaires (2.3%). Other research provides similar evidence of the increased reporting of socially undesirable and high-risk behaviors among participants who provide data with the CASI method.[6,8,12,17]

Limitations

CASI and audio-CASI also have limitations. The technology is costly: the researcher must purchase laptop computers, high-speed Internet connections, and Internet routers for data collection. The researcher may also have to hire a computer programmer to write the type of program needed to present the survey questions in a manner appropriate for children and for exporting responses to the secure Web site for statistical analysis at a later time.

When research instruments need to be translated into other languages, additional expenses are incurred when these items are written for the CASI format and when the audio format is added for those who need the items to be read to them. In addition to the high cost of start-up, the technology is subject to failures such as a laptop computer that fails to respond correctly or an Internet connection that is down for a period of time.

APPLICATION IN THE FIRST WAVE OF A LONGITUDINAL STUDY

The administrators of 3 school districts in central Texas reviewed and approved the proposed study and agreed to participate. The study was reviewed and approved by the University's Institutional Review Board.

During initial planning for data collection, school administrators were positive about the opportunity for their participating students to use computer technology. While all of the schools had some computers available in a lab or library for small class exercises, individual children had limited time using computers. The participating schools had 35% to 75% of their students eligible for free or reduced price lunch, which was an indicator of economic disadvantage in these schools. Therefore, it was inferred that many students had little or no computer experience at home.

Pilot study with children

Prior to using the audio-CASI method in the first wave of a longitudinal study, a pilot study was conducted with 9 Hispanic children (4 males and 5 females) who attended one of the schools in which the study was to take place. Children were given the choice of completing the survey using the paper-and-pencil method or using a laptop computer. Seven of the 9 children chose the computer option. Their qualitative assessments of the experience included comments such as "it was fun" and "it was easy." Furthermore, there was no difference in the amount of time needed to complete the survey by CASI than by the paper-and-pencil method.

Longitudinal sample

The sample consisted of 1161 children in grades 4 to 6. There were 45% boys and 55% girls, with 50% Latino, 36% White, 12% African American, and 2% other ethnic group members. Information was sent home to the parents of all children in grades 4, 5, and 6 and included a cover letter explaining the study and parent consent forms. The consent forms were to be returned to the schools for collection by the research team.

Data collection protocol

A schedule for data collection was arranged with each school. Once on site the research team would sign in at the school office and then proceed to the classroom where data collection was to take place. The classroom was prepared by setting up the laptop computers and establishing the wireless Internet connection. In a few instances, data collection took place in the school's computer lab and so the school's computers were connected to the secure Web site as well and used as additional data collection stations. The time needed to set up the computer equipment was 10 to 15 minutes, but this task was performed only once each day.

At the start of a data collection session, the participating children entered the classroom and were assisted in finding a computer station. The study was explained, questions answered, and written assents obtained for each participating child. Children were given the option to complete the study in either English- or Spanish-language versions. Special education students and those with reading skills below grade level were equipped with earphones for the audio-CASI that was also available in English and Spanish. No children declined to participate owing to reading problems. The schools allocated one full hour for the participants to complete surveys. Nearly all fifth and sixth graders and the majority of fourth graders completed the entire survey in the allotted time. Make-up sessions were scheduled for those who did not complete the survey in one hour.

After the research staff explained the study and children were seated at the appropriate computer (ie, English/Spanish, audio/nonaudio), the research staff logged each participant onto the computer and gave them instructions to start the survey measures. At the end of each section in the survey the child was instructed to "send" the answers. In this way sections of the completed survey were routed to the secure Web site and compiled in a Microsoft Access Database for analysis at a later time. The computer program tracked when each participant started the survey and when he/she finished, and calculated how long it took to complete the survey. There were 2 advantages to transmitting sections of the survey. First, the Internet router was not overloaded with the simultaneous

transmission of the full survey for the entire group. Rather, smaller packets of information were transmitted at varied times as the individual children completed each section at their own pace. Second, in the occasional case of a computer shutting off (eg, when the battery was low, the computer went into the energy saver mode), the sections that the child had already sent to the secure Web site were safely in the data bank. When this happened, the child was given the option of completing the remaining survey by the paper-and-pencil method or moving to a new unused laptop computer.

Data analysis procedure

Data collected via CASI was routed directly to a secure Microsoft Access Database. From there the research staff was able to manipulate the data and prepare it for analysis. The data were reviewed for format and cleaned as needed.[5] The data bank could then be uploaded into a variety of programs for further statistical cleaning and analysis. In this longitudinal study, the data were imported from the Microsoft Access Database into Microsoft Excel. This was done to allow the research team to do data entry on the small number of paper-and-pencil surveys that were collected from children who missed the scheduled days for data collection or who did not have time to complete all of the survey at the scheduled time. After the data entry was complete the data set was then imported into SPSS for statistical analysis, saving hours of staff time in manually entering data and eliminating data-entry errors.

Participant benefits

In addition to the benefits of using CASI methods described above, in longitudinal studies the storage of data is managed electronically rather than through warehousing thousands of paper copies of surveys. While CASI is beneficial to the researcher who is conducting a large-scale study, it also has benefits for the participants. School-age children spend the majority of their school day working with paper-and-pencil activities. The opportunity to work at the computer provided a welcome change to the usual school-day routine. Consequently, the children's attention to this task was held longer. The computer provided a more interactive format, and the children perceived their work on the computer to be more like that of a game, and less like an exam.[20] In addition, the computer program allowed for the use of customized graphics on the computer screen, which is not possible in a paper survey, and this also helped to keep the children's attention engaged. The children received positive feedback from the computer program as they completed the different sections.

DEVELOPMENT AND IMPLEMENTATION CHALLENGES OF CASI

Developing a computer-assisted survey presents a number of challenges, both technical and nontechnical. The software developer must be cognizant of these challenges to meet the needs of the research team.

Nontechnical challenges

The nontechnical challenges refer to the way the software system functions for the different users. The primary nontechnical challenge of software development for CASI is to understand the system requirements for meeting the researchers' needs. The software system selected should (*a*) provide the research team with a means of entering questions into the system for accessing at a later time; (*b*) provide participants with a means for viewing or hearing the survey questions and responding to them; and (*c*) enable the researchers to analyze the survey data.

Entering questions

One challenge before developing the system is to determine who will be the survey author, or the person who enters questions into the system, and what level of computer knowledge and skills competencies he/she should possess. The system's means of entering survey questions must be easy enough

to accommodate the author's computer skills. The author not only enters questions, but also indicates to the system in which format to store the questions and answer choices for future analysis. With a multiple-choice question, for example, the author associates each possible answer with a numerical response. To keep the attention of survey respondents, the author may also add pictures, insert words of encouragement, or vary the color of the background on which the survey text appears.[20] If the respondents have limited literacy, the author may include audio recordings of the instructions, questions, and answer choices that the survey respondents can play.

Viewing and responding to questions

During data collection, the computer interface must be rich enough to allow for the inclusion of pictures, words of encouragement, and color variations throughout the survey. The participants must be able to indicate an answer response to the questions. If fill-in-the-blank or essay questions are included in the survey, the participants must be able to enter text. The software system should make it easy for participants with little computer experience to complete the survey.

Analyzing data

The software system should enable the researchers to analyze the data obtained from participants. Rather than requiring the research team to consolidate participants' responses from potentially thousands of different computers, the system should allow the researchers to organize, manipulate, graph, and perform calculations on the data stored in the secure Web site. The researchers should not require additional technical computer skills beyond those of an experienced Excel or SPSS user.

Technical challenges

The primary technical challenges in developing a computer-assisted survey include (*a*) choosing networking technologies; (*b*) deciding whether to buy versus build certain pieces of the survey system; and (*c*) making large amounts of data available for analysis.

Networking technology

The chosen networking technologies determine whether the system will accommodate the number of simultaneous participants while avoiding complex hardware set-up. The technical requirements for this study included setting up an 802.11x wireless network in the classroom to enable the research team to connect dozens of participants to the Internet simultaneously. Each of the laptop computers used in this study had an 802.11x card that enabled the computer to connect to the secure Web site.

If the program developer publishes the survey on the World Wide Web in a secure site, the participants can complete surveys using a Web browser, which comes preloaded on virtually every computer purchased. The World Wide Web, based primarily on HTML, presents a standard and familiar user interface to most of the participants. Respondents require little instruction or training to understand how to read, answer, and submit questions that appear in a Web browser. Although simple, an HTML-based interface is rich enough to enable the system to include graphics and sounds, and vary the background colors of pages, thereby maintaining the attention of survey respondents. It also includes in its standard set of user interface elements buttons, radio buttons, checkboxes, and text boxes, which enable the system to display multiple choice, multiple response, essay, and fill-in-the-blank questions. Finally, the secure Web server can record the participants' responses in a central database that the researcher can retrieve at a later time.

Survey system

The program developer can use an existing survey development software product or build the survey system from scratch. Building it from scratch yields a highly customizable system, but increases the amount of time and, therefore, the cost of developing the

program. Buying an existing survey development software package, such as Question Mark's Perception,[21] not only decreases development time, but may also have the added benefit of being familiar to members of the research team.

Data availability for analysis

The simplest way of making large amounts of data available for analysis is to provide a facility for importing data into a standard analysis tool such as Excel or SPSS. The research team requires little training to use these tools with which they are already familiar. In this study the data had been stored on the secure Web site in a relational database. In this format, the research team can use the standard import functions in the statistical program to load the data and begin analyzing it. The program developer provides the research team with the database query scripts to load the data the researchers want to analyze.

IMPLICATIONS FOR FURTHER STUDY

The application of computer technology expands the possibilities of conducting increasingly more complex studies with larger numbers of participants. The use of this technology in obtaining qualitative data has not been explored but could eliminate the costly and time-consuming processes of training and employing interviewers, including those who are bilingual, and transcribing qualitative data. Special programs could be written that include follow-up questions to probe for additional qualitative data. The Internet also presents further opportunities for wide-ranging and flexible data collection.[22]

CONCLUSIONS

CASI is a promising use of information technologies for further research. Although expensive to start-up, it has clear financial advantages for longitudinal studies. Moreover, the advantages of obtaining more reliable and valid data than with traditional paper-and-pencil methods or person-to-person interviewing offset the financial challenge. Using CASI with children presents a more interesting format for obtaining large amounts of data while providing children with a constructive experience using the computer.

REFERENCES

1. United States Department of Health and Human Services. *Healthy People 2010*. McLean, Va: International Medical Publishing Inc; 2000.
2. Brener ND, Collins JL. Co-occurrence of health-risk behaviors among adolescents in the United States. *J Adolesc Health.* 1998;22:209–213.
3. Kann L, Warren CW, Harris WA, et al. Youth risk behavior surveillance—United States, 1993. *J Sch Health.* 1995;65:162–171.
4. Rew L, Horner SD. Resilience framework for reducing health-risk behaviors in youth. *J Pediatr Nurs.* 2003;18:379–388.
5. Knapp TR. *Quantitative Nursing Research.* Thousand Oaks, Calif: Sage; 1998.
6. Metzger DS, Koblin B, Turner C, et al. Randomized controlled trial of audio computer-assisted self-interviewing: utility and acceptability in longitudinal studies. *Am J Epidemiol*. 2000;152:99–106.
7. Aquilino WS, Wright DL, Supple AJ. Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use Misuse*. 2000;35:845–867.
8. Ellen JM, Gurvey JE, Pasch L, et al. A randomized comparison of A-CASI and phone interviews to assess STD/HIV-related risk behaviors in teens. *J Adolesc Health*. 2002;31:26–30.
9. Millstein SG. Acceptability and reliability of sensitive information collected via computer interview. *Educ Psychol Measurement*. 1987;47:523–533.
10. Turner CF, Ku L, Rogers SM, et al. Adolescent sexual behavior, drug use, and violence: increased computer survey technology. *Science*. 1998;280:867–873.
11. Williams ML, Freeman RC, Bowen AM, et al. A comparison of the reliability of self-reported drug use and sexual behaviors using computer-assisted versus face-to-face interviewing. *AIDS Educ Prev*. 2000;12:199–213.
12. Newman JC, Des Jarlais DC, Turner CF, et al. The differential effects of face-to-face and computer

interview modes. *Am J Public Health*. 2002;92:294–297.
13. Mensch BS, Hewett PC, Erulkar AS. The reporting of sensitive behavior by adolescents: a methodological experiment in Kenya. *Demography*. 2003;40:247–268.
14. Norusis MJ. *SPSS® 8.0 Guide to Data Analysis.* Upper Saddle River, NJ: Prentice-Hall Inc; 1998.
15. Supple AJ, Aquilino WS, Wright DL. Collecting sensitive self-report data with laptop computers: impact on the response tendencies of adolescents in a home interview. *J Res Adolesc*. 1999;9:467–488.
16. Van Griensven F, Supawitkul S, Kilmarx PH, et al. Rapid assessment of sexual behavior, drug use, human immunodeficiency virus and sexually transmitted diseases in Northern Thai youth using audio-computer-assisted self-interviewing and noninvasive specimen collection. *Pediatrics.* 2001;108(1):e13. Accessed September 8, 2003.
17. Riley ED, Chaisson RE, Robnett TJ, et al. Use of audio computer-assisted self-interviews to assess tuberculosis-related risk behaviors. *Am J Crit Care Med*. 2001;164:82–85.
18. Dixon SD, Stein MT. *Encounters With Children: Pediatric Behavior and Development.* St. Louis: Mosby; 2000.
19. Lessler JT, O'Reilly JM. Mode of interview and reporting of sensitive ssues: design and implementation of audio computer-assisted self-interviewing. *NIDA Res Monogr*. 1997;167:366–382.
20. Graue ME, Walsh DJ. Studying children in context: theories, methods, and ethics. Thousand Oaks, Calif: Sage. 1998.
21. Question mark's perception. Available at: http://www.questionmark.com/us/perception/index.htm. Accessed September 29, 2003.
22. Im E-O, Chee W. A feminist critique on the use of the internet in nursing research. *Adv Nurs Sci.* 2001;23(4):67–82.