

# TECHNICAL PAPERS ON HEALTH AND BEHAVIOR MEASUREMENT

## TECHNICAL PAPER 11

### **Audio and Video Computer-Assisted Self-Interviewing: Preliminary Tests of New Technologies for Data Collection**

James M. O'Reilly, Michael L. Hubbard, Judith T. Lessler,  
Paul P. Biemer, and Charles F. Turner

#### **Reference Citation**

O'Reilly, J.M., M.L. Hubbard, J.T. Lessler, P.P. Biemer, and C.F. Turner. (1994) Audio and video computer-assisted self-interviewing: Preliminary tests of new technologies for data collection. *Journal of Official Statistics* 10(2):197-214.

# Audio and Video Computer Assisted Self-Interviewing: Preliminary Tests of New Technologies for Data Collection

*James M. O'Reilly<sup>1</sup>, Michael L. Hubbard<sup>2</sup>, Judith T. Lessler<sup>1</sup>, Paul P. Biemer<sup>2</sup>, and Charles F. Turner<sup>3</sup>*

**Abstract:** Results are reported from a preliminary study testing a new technology for survey data collection: audio computer assisted self-interviewing. This technology has the theoretical potential of providing privacy (or anonymity) of response equivalent to that of paper self-administered questionnaires (SAQs). In addition, it could offer the advantages common to all computer assisted methods such as the ability to implement complex questionnaire logic, consistency checking, etc. In contrast to Video-CASI, Audio-CASI offers these potential advantages without limiting data collection to the literate segment of the population. In this preliminary study, results obtained using RTI's Audio-CASI system were compared to those for paper SAQs and for Video-CASI. Survey questionnaires asking about drug use, sexual behavior, income, and demographic characteristics were administered to a small sample ( $N = 40$ ) of subjects of average and below-average reading abilities using each method of data collection. While the small

sample size renders many results suggestive rather than definitive, the study did demonstrate that both Audio- and Video-CASI systems work well even with subjects who do not have extensive familiarity with computers. Indeed, respondents preferred the Audio- and Video-CASI to paper SAQs. The computerized systems also eliminated errors in execution of "skip" instructions that occurred when subjects completed paper SAQs. In a number of instances, the computerized systems also appeared to encourage more complete reporting of sensitive behaviors such as use of illicit drugs. Among the two CASI systems, respondents rated Audio-CASI more favorably than Video-CASI in terms of interest, ease of use, and overall preference.

**Key words:** Survey measurement; survey technology; computer assisted interviewing (CAI); Audio-CASI; Video-CASI; self-administered questionnaires; sensitive behaviors; sexual behavior measurement; drug use measurement; income measurement.

<sup>1</sup> Battelle Memorial Institute, 100 Capitola Drive, Durham, NC 27713, U.S.A.

<sup>2</sup> Research Triangle Institute, P.O. Box 12194, Research Triangle Park, NC 27709, U.S.A.

<sup>3</sup> Research Triangle Institute, 6101 Executive Boulevard, Rockville, MD 20852, U.S.A.

**Acknowledgements:** The research reported herein and

preparation of the original draft of this article were supported by the Research Triangle Institute. Revision of this manuscript for publication was supported, in part, by grants 1-R01-HD31067-01 and 5-P50-DA06990-02 (Measurement Error Component) from the National Institutes of Health to Charles Turner.

## 1. Introduction

The written self-administered questionnaire (SAQ) is a standard technique used for in-person surveys to question respondents on drug use, sexual behaviors, and other sensitive matters. The written SAQ provides respondents with privacy in answering sensitive questions. The written SAQ, however, imposes substantial limitations on data collection. It requires literate respondents, and it limits the extent to which contingent questioning strategies (using "skip" instructions) can be employed.

Some of the restrictions imposed by the written SAQ can be overcome by the use of a computer to present questions. Video computer assisted self-interviewing (Video-CASI) has been used occasionally since at least the 1960s (see, for example, Evan and Miller 1969). Technologically, Video-CASI may be considered a variant of the widely used computer assisted personal interviewing (CAPI) technology. It differs from CAPI in that the respondents themselves – rather than the interviewers – interact with the computer. Video-CASI technology facilitates use of complex questionnaire logic and branching, since these are performed by the computer and do not burden the respondent with the need to follow complex instructions. Video-CASI does not, however, eliminate the requirement of literacy. As with written SAQs, Video-CASI presumes that respondents are sufficiently literate to read the survey questions and select a response.

Audio computer assisted self-interviewing (Audio-CASI), as developed at the Research Triangle Institute (RTI), offers the promise of eliminating the requirement of literacy that limits both written SAQs and Video-CASI. At the same time, Audio-CASI provides the major advantage

common to all self-administered modes – greater respondent privacy – plus the unique advantages of computerized administration, e.g., exactly standardized administration of the questions, instantaneous range and consistency checking of responses, convenient multilingual administration, etc.

In an Audio-CASI interview, the computer plays a recorded version of questions and answer choices over headphones, and the subject responds through the keyboard. The computer records the response and, based on it, plays the next appropriate question. Technologically, the key element of Audio-CASI is that questions are presented as good quality voice recordings that are stored in digital form on the PC's disk. Questions are accessible in random fashion, allowing complex question ordering and therefore detailed questioning of those reporting certain behaviors. Earlier audio technologies such as tape do not allow such rapid random accessing of items (see, for example, Camburn, Cynamon, and Harel 1991).

We briefly review below issues that arise in surveying the public on sensitive issues and provide details of RTI's Audio-CASI hardware and software systems. We then discuss our initial test of the efficacy and acceptability of the system on a small sample of average and below-average readers.

### 1.1. Sensitive questions and self-administration

The researcher who needs to ask people questions about sensitive behaviors is faced with the need to balance several competing research goals. He or she would like to encourage the respondent to provide complete and truthful responses. At the same time, the researcher would like to

gather detailed information about the sensitive behavior and to have the ability to characterize the people who may report a sensitive behavior.

Although the results have not always been strong or consistent across studies (e.g., Bradburn and Sudman 1979, p. 8–13), it is generally concluded that this research has shown that more private methods – self-administered forms and randomized response – produce more valid reports of sensitive behavior (Bradburn 1983; Miller, Turner, and Moses 1990, ch. 6; Schwarz, Strack, Hippler, and Bishop 1991). However, these more private questioning methods have disadvantages that often interfere with other research goals. Randomized response and similar techniques<sup>4</sup> permit aggregate estimates of the prevalence of sensitive behaviors; however, they are severely limited in that the researcher does not know the correct answer for each individual in the study. So it is impossible to characterize people with and without the sensitive behavior. In addition, neither of these techniques allows one to ask detailed follow-up questions about the sensitive behavior.

Thus, many studies make use of self-administered forms to gather information on sensitive issues. Several recent studies have confirmed the superiority of self-administered techniques for gathering sensitive information. Hay (1990), for example, reports on a study in which a sample of 1,502 students in grades 6 through 12 were randomly assigned to receive either a self-administered questionnaire or a personal interview. Seventy-four percent of those answering with the self-administered form versus 63% of those answering the

personal interview reported ever drinking an alcoholic beverage. The analogous comparison for use of cigarettes was 38% versus 30%.

Turner, Lessler, and Devore (1992) report similar findings for a wide range of age groups reporting on their use of a variety of licit and illicit drugs. In their large-scale field test of alternative data collection methods for the National Household Survey on Drug Abuse, some 3,200 respondents were randomly assigned to receive either a self-administered or an interviewer-administered version of the questionnaire.

Table 1 presents the ratio of reported drug use for the self-administered versus interviewer-administered versions. Results for three drugs that vary in sensitivity are shown – cocaine, marijuana, and alcohol, and the ratios are shown for three time periods – the 30-days prior to the interview, the 12-month period prior to the date of interview, and at any time during the person's lifetime.

The results consistently indicate that as the survey questions become more socially sensitive, the interviewer-administered format produces lower reports. So, for example, Turner et al. (1992) found that the proportion of people reporting cocaine use in the 30 days prior to the date of interview was 2.4 times greater when measurements were made in the self-administered format than when the interviewer asked the questions. In contrast, for alcohol use the corresponding ratio is only 1.08. The ratio for marijuana use falls between these two at 1.61. Although there were few overall differences in responses by mode for alcohol use, Turner et al. (1992, p. 188, Fig. 7-2) report finding significant differences for younger cohorts of respondents. Thus respondents 12 to 17 years of age were 1.38 times more likely to report

<sup>4</sup> Similar techniques include item count (Droitcour et al. 1991), and scrambled randomized response (Ahsanullah and Eichorn 1984). See Umesh and Peterson (1991) for a review of randomized response techniques.

*Table 1. Ratio of estimates of prevalence of drug use obtained in self-administered versus interviewer-administered questionnaires*

	Lifetime	Last 12 months	Last 30 days
Alcohol	0.99	1.04	1.06
Marijuana	1.05	1.30	1.61
Cocaine	1.06	1.58	2.40

*Note.* Ratios are weighted proportions of respondents reporting use of drug in self-administered questionnaire (SAQ) divided by proportion of respondents reporting use of drug in interviewer-administered questionnaire (IAQ) format. Estimates of proportions are derived from survey of 3,326 respondents who were randomly assigned to either IAQ or SAQ conditions. Respondents were a probability sample of persons 12 years of age and older residing in 33 purposely-sampled metropolitan areas in the U.S.A. in 1990.

Source: Turner et al. (1992, Fig. 7-1).

alcohol use in the past 30 days in the self-administered format.

Jones and Forrest (1992) report similar results for the use of self-administered questionnaires to improve the reporting of abortions in the National Survey of Family Growth. They compared the estimated number of abortions from the survey to external national estimates. They found that without the self-administered form, the survey estimate was only some 39% of the external estimate. Using the self-administered form, the survey estimate was about 71% of the external estimate.

While recent evidence increasingly points to the use of self-administered forms as superior for asking about sensitive behaviors, there are difficulties with this approach. Most importantly is that substantial proportions of the population possess limited literacy (National Center for Education Statistics 1993). In addition, even people who can read well may be limited in "forms literacy," that is, in the

ability to understand and follow the conventions of data collection forms (Lessler and Holt 1987).

Because of these difficulties, researchers have been forced to limit their use of complex questioning strategies when designing self-administered forms. Typically, the number of questions is restricted, and the use of contingent questioning is minimized or eliminated.

### *1.2. Experiences with Video-CASI and CAPI*

Computerized self-administration of questionnaires provides an alternative that mitigates the burdens imposed by complex branching schemes in questionnaires. Since at least the late 1960s, researchers have reported attempts to use computers – initially mainframe and minicomputers – for the self-administration of questionnaires. The conclusions of this research on self-administration of questionnaires parallel in some respects the conclusions of research on the computerization of interviewers tasks in telephone (CATI) and personal (CAPI) interviewing. Thus the earliest reported study of Video-CASI (Evan and Miller 1969) provided evidence that college undergraduates placed somewhat greater trust in the anonymity of the responses they provide in a Video-CASI versus paper SAQ format. This was accompanied by more complete reporting of sensitive data (e.g., reporting of anxieties, feelings of anomie, etc.).

The *potential* operational advantages of computer assisted interviewing for interviewer-administered surveys has long been recognized. As with CASI, these potential advantages include the ability to employ exceedingly complex questionnaire logic, customized "fills" in questions, automated prompting and consistency checking, etc.

(Weeks 1992). Empirical evidence of the extent to which CATI and CAPI systems have converted these *potentials* into significant reductions in survey measurement errors is, however, more limited (Groves and Nicholls 1986; Weeks 1992). There is also a growing recognition that computer assisted interviews can introduce their own unique sources of error such as those introduced by undetected programming errors.<sup>5</sup>

Nonetheless, the available empirical evidence indicates that correctly functioning CAPI and CATI systems are well accepted by both interviewers and respondents (Weeks 1992; Baker and Bradburn 1991; Bradburn et al. 1993). Furthermore, there is some, albeit inconsistent, evidence suggesting that use of a computer produced slight increases in the reporting of socially undesirable behaviors. Thus Baker and Bradburn (1991) report finding two instances in Round 11 of the National Longitudinal Survey of Youth Cohort (NLS-Y) in which there was significantly more frequent reporting of alcohol use when interviewers used CAPI rather than a paper and pencil questionnaire. However, evidence from further tests built into Round 12 of the NLS-Y revealed small but significant increases as well as decreases in the reporting of socially desirable behaviors. Thus more males reported using birth control during the previous month when questioned using CAPI (66% versus 58%) but among all males who

reported using birth control during the preceding month, CAPI respondents reported consistent use of birth control less frequently (89% versus 93%).

### 1.3. Overview

Researchers, then, have been faced with the dilemma of needing private, self-administered methods to question respondents effectively on sensitive topics. Yet the written SAQ required us to use very simple questioning strategies and both the written SAQ and Video-CASI formats could not obtain data from the sizable segment of the population with limited reading skills.

The present article reports results from preliminary testing of our new Audio-CASI technology, and it compares these results with those obtained using Video-CASI and written SAQs. As an exploratory study, the sample sizes for our research were small, and thus our findings will often be suggestive rather than definitive. Nonetheless, we hope readers will find important suggestions concerning both the potential value of our new Audio-CASI technology as well as some unsuspected arguments for the earlier – and, we believe, under-utilized – Video-CASI technology.

Before turning to details of our research design and results, we present below a brief description of the hardware and software used to construct RTI's Audio-CASI system.

## 2. Audio-CASI System

In 1991, as an outgrowth of RTI's work on the National Household Seroprevalence Survey, the National Household Survey on Drug Abuse, and other projects studying sensitive behaviors, researchers at RTI began investigating alternative technologies for collecting data on sensitive

<sup>5</sup> What is "new" about errors in CASIC surveys is that mistakes by programmers (for example, overwriting of prior data fields) can introduce inadvertent data loss and other systematic errors. Obviously, good programming practices and better designed software tools would detect most such errors prior to the fielding of CASIC surveys but we suspect, that some errors will occasionally find their way into even the most careful work. Paper and pencil surveys may be vulnerable to similar errors in data entry programs but these can usually be remedied by reprocessing the forms. Programming errors in CASIC surveys often can be remedied only by repeating part of the data collection.

topics.<sup>6</sup> We began by specifying a number of key requirements to guide the development. These were:

- Good quality recorded audio, not synthesized sound;
- No significant delays in playback of audio;
- A robust system usable by average field interviewers in a broad range of environments; based on light-weight, economical laptop systems available today;
- Audio software integrated with a standard computer assisted interviewing (CAI) software system<sup>7</sup> – not a new, custom CAI application;
- Broad flexibility in administration, permitting respondents to take all or part of interview with audio alone, with video alone, or mixed.

By late 1991 a prototype system had been developed and testing had begun (O'Reilly, Lessler, and Turner 1992a, 1992b; O'Reilly and Turner 1992). An MS-DOS system was selected as the hardware platform rather than the Apple Macintosh because of hardware prices, the availability of a number of widely used CAI software systems for MS-DOS, and the emergence of portable MS-DOS audio-digital devices. The key hardware components of the 1991 RTI prototype system were:

- a laptop PC weighing six pounds and running DOS 5.0 with: a 16 MHz, 80286 processor; 1 Mbyte RAM;

<sup>6</sup> A parallel effort was also underway in 1991 at the University of Michigan using the Apple Macintosh as a hardware platform (Johnston 1992).

<sup>7</sup> Computer assisted interviewing (CAI) systems include CAPI, CASI and CATI applications. Except for the amendments that may be required to accommodate specialized input or output devices and the needs of "novice" computer users (rather than professional survey interviewers), CAPI and CASI systems are often virtually identical.

60 Mbyte hard disk; a VGA monochrome display; and

- a digital audio device interacting with the laptop PC through the parallel port.

### 2.1. Software

The base CAI software is the Blaise system developed by Statistics Netherlands (1989) and widely used in European government statistical agencies. The audio device used an RTI-developed driver to interact between the CAI questionnaire and the audio digital device. (The RTI driver can function equally well with other CAI packages besides Blaise.) Audio functions are external to the CAI system, and no changes were made to the normal mode of operation of Blaise in order to add audio.

The RTI Audio-CASI system implements a number of supporting capabilities through the PC function keys: display or blanking of questions on the screen; turning audio on or off, repeating the question, backing up, and audio "help." Other features of computer-assisted interviewing such as range and consistency checks are also implemented in audio mode. Audio feedback of the selected response is another important feature. This confirms for the respondent that the correct choice has been entered.

### 2.2. CAI authoring system

To implement Audio-CASI for operational use, RTI developed a CAI software authoring system. Audio-CASI adds an extra layer of complexity to the already challenging task of programming a sturdy CAI instrument. Each question has an associated DOS audio file and usually another audio file for the answer choices. The syntax to permit the audio echoing of

the selected response adds another level of complexity.

This complexity means that modifying and adapting a CAI instrument of substantial size would be difficult, time consuming, and problematic. RTI has developed a CAI authoring system in the FoxPro-2 relational database language. The question components (name, text, answer choices, and associated audio files) are defined in database records, and the questionnaire logic specified in a related table. The CAI authoring system generates the CAI language application code. With this generated code the Blaise system compiles and generates the Audio-CASI instrument. The authoring system makes modifying CAI instruments much less difficult. A new version of the CAI instrument source code, reflecting the changes, can be generated and tested in a fraction of the time of traditional methods.

### 3. Research Design

For this first investigation of data collection using Audio-CASI technology, we compared three modes of self-administered data collection: paper and pencil SAQ, computer assisted self-interviewing with respondents reading questions on the PC screen (Video-CASI), and our Audio-CASI system (as described above). This test was conducted with a small sample ( $N = 40$ ) of average and weak readers. Three different questionnaires on income, drug use, and sexual experience were administered (one in each mode). Then a brief interview was conducted asking the respondents' preferences among the three modes.

This is properly understood as an exploratory study. We had four major aims, two of which were within the aspirations of such a preliminary study:

1. Testing the feasibility of using an Audio-CASI system for data collection;
2. Assessing subjects' reactions to Audio-CASI interviewing (in comparison to traditional paper SAQs and Video-CASI).

More ambitious goals of this research were inevitably compromised by the exploratory nature of this study. Due to our small sample size, it will only be possible to adduce *suggestive* evidence on two of our other goals:

3. Assessing the effects of Audio-CASI on the technical quality of the data collected (e.g., accuracy of "branching," extent of item nonresponse).
4. Assessing the effects of Audio-CASI on the willingness of subjects to report sensitive behaviors (e.g., drug use, certain sexual behaviors).

#### 3.1. Experimental design

Our experimental design has one between-subjects factor – the reading level of the subjects (average versus below average). Within-subjects variables were mode of question administration and content of questions. The order of presentation of modes and contents was completely counterbalanced between subjects. In addition, all combinations of mode and content were produced using a greco-latin square experimental design.

#### 3.2. Subjects

Subjects in the experiment were forty adults who participated as paid volunteers. Thirty-five of the 40 subjects were recruited at a local community college. Average reading-level subjects ( $N = 15$ ) were students in a course for completion of the General

Education Degree (GED).<sup>8</sup> They were approximately ninth to twelfth grade readers. Below average reading-level subjects ( $N = 20$ ) were students in pre-GED classes. Their reading levels ranged from fifth through eighth grade level. An additional five subjects (included as average reading level subjects) were recruited from a list of volunteers maintained by RTI's Laboratory for Survey Methods and Measurement. All had educations ranging from high school to some college.

Subjects ranged in age from 17 to 49.<sup>9</sup> There were 5 males and 31 females; 20 blacks, 15 whites, and 1 American Indian; 10 subjects were employed and 26 were not employed. Four subjects were married, 4 were separated, 4 were divorced, 2 were widowed, and 22 were never married.

### 3.3. Testing procedures

Testing of subjects was conducted at two locations. The 35 community college subjects were interviewed in a library-resource room at the community college. Interviews were conducted in corners of a large room where other people, tables, books, and audio-visual equipment were also located. While the privacy of the interviews was maintained, it was possible that someone could inadvertently pass nearby and see or overhear the interview. The five other subjects were interviewed at RTI's Laboratory for Survey Methods and Measurement. Interviews were conducted in small testing rooms where only the interviewer and the subject were present, and there

was no possibility of being overheard by anyone else.

After the nature of the interview was explained, subjects completed the three sections of the interview designated by the experimental design. While subjects were completing each section, the interviewer timed each of the three parts of the interview, and also observed the subject as he or she completed the self-administered interviews.

Interviewers recorded the number of requests for assistance by the subject, a general rating of how easy it was for subjects to read the various parts of the interview, and general ratings of how difficult it was for subjects to use the computer. Following completion of the three self-administered instruments, the interviewer asked subjects to give their general impressions of the three modes of questioning and they asked eight questions about the subjects' perceptions and preferences among the three modes of interviewing. These comments were recorded and later transcribed.

Subjects were asked to rank the three modes of questioning in terms of:

- Which mode affords the most privacy while answering questions;
- Which mode best protects the privacy of their answers once the interview is completed;
- Which mode most encourages honest and truthful answers;
- Which mode is the easiest to use;
- Which mode is most interesting to use;
- Which mode made it easiest to change answers;
- Which mode is best for answering sensitive questions like the ones they just completed;
- Which mode they like most.

Following this debriefing interview, subjects

<sup>8</sup> The GED is a diploma in the U.S.A. signifying completion of education equivalent to secondary school, i.e., 12th grade. Reading levels were described by the staff of the GED program.

<sup>9</sup> Demographic information was missing for four respondents.

signed permission forms allowing interviewers to use the recordings of the debriefing interview for research purposes. Then they were paid for their participation and dismissed.

### 3.4. Question content

There were four different sets of questions concerning background demographic characteristics, drug use, income, and sexual behavior. The background demographic questions were always administered by CASI (both with and without voice) as an introduction to using the computer. The other three question areas were prepared in paper SAQ, Video-CASI, and Audio-CASI formats, so that the same question content could be presented in all three modes. All sets of questions contained branching, where the next question to answer was conditional on the answer to a previous question. In the paper SAQ, the subject followed an arrow to either the next question or to an instruction box, printed in a shaded box, which indicated the next question to answer. For the computerized questioning modes, these branches were controlled by the computer.

Background demographic questions asked subjects to report their gender, age, date of birth, employment status, marital status, the number of persons living in the subject's household, their relationship to other household members, Hispanic ethnicity, and race.

Drug use questions asked about the subject's use of cigarettes, alcohol, marijuana, cocaine, and crack. All subjects answered yes-no questions about use within the past 30 days, use within the past 12 months, and lifetime use for each drug. When subjects indicated that they had used any drug within one of these three reference periods,

they were asked additional questions:

- for past 30 days: days used in the past 30 days, amount used in past 30 days, frequency of use in the past 12 months, and age at first use;
- for past 12 months: frequency of use in the past 12 months and age at first use; and
- for lifetime drug use: age at first use of cigarettes and alcohol, age at first use plus total lifetime days used for marijuana-hashish, cocaine, and crack.

Income questions asked about total income in the past calendar year, followed by specific questions about income from various sources (wages or salary, self-employed income, farming income, income from interest, dividends, or royalties, income from social security, Supplemental Security income, income from welfare or public assistance, income from child support, and income from other sources). Additional questions were asked about making child support payments, missing child support payments, amount of charitable donations, income not reported on income tax reports, credit card ownership and credit limits, and exceeding credit card limits. For all but the total income and the charitable donation questions, subjects were first asked whether or not they had received a given source of income in the past year. If they had, they were directed to a question about the amount of that income; if they had not, they were asked about income from the next source. Similar branching was done for questions about nonpayment of child support, unreported income, and credit card limits.

Sexual behavior questions were different for men and women. All men were asked identical sets of questions about heterosexual and homosexual behavior. Women were asked only about heterosexual

behavior and pregnancy outcomes. Sexual behavior questions included questions about types of sex acts, sex with different types of partners (current steady partner, IV drug users, bisexual partners, pick-ups or casual dates, and prostitutes (men only), being forced to have sex (women) or forcing someone to have sex (men), age at first sex, and number of lifetime partners. In addition, the following questions were asked about the subject's first three sexual partners: recency of sex, time since first met partner, interval between first meeting and first sex, number of times had sex, frequency of condom use with partner, and whether protection from STDs and birth control were discussed with the partner. Pregnancy outcome questions asked about: ever being pregnant (with "skip" of subsequent questions if never pregnant), number of pregnancies, number of live births, number of miscarriages or stillbirths, and number of abortions.

## 4. Results

### 4.1. PC operations

Very few subjects had used computers more than a few times prior to participating in this research. Yet all but one or two subjects became comfortable using the screen and keyboard after a few minutes of instruction. The interviewers were available for support and assistance throughout the sessions but after the first few minutes they were rarely called on to help.

### 4.2. Audio-CASI operations

Overall the system worked effectively and reliably. Subjects were able to administer the audio questionnaires to themselves with little difficulty. Questions and requests

for assistance were no more frequent than for the paper self-administered questionnaire (SAQ).

### 4.3. Statistical analyses

#### 4.3.1. Subject reactions

Results are presented in Table 2 for the ranking of the three modes of questioning, the interviewer-recorded response times, the number of requests for help, and the proportion of skip instructions correctly followed. These results were analyzed using a repeated measures analysis of variance. The between-subjects factor was the reading level of the subjects. The within-subjects factor was mode of presentation. The analyses reported in Table 2 are mean rankings by mode of interview ignoring content and reading level. The  $p$ -values reported in Table 2 represent the statistical significance of the mode effect relative to the appropriate error term (where effects of reading level,<sup>10</sup> subjects within reading level, mode and reading level  $\times$  mode interaction are removed from the error term). Additional tests of differences among the three modes were made using  $t$ -tests.

The two computer assisted modes of data collection (Audio- and Video-CASI) were judged superior to paper SAQs (with  $p < 0.05$ ) on eight of the nine rating scales (liked best, ease of use, most interesting to use, best for asking sensitive questions, best for getting honest answers, best for privacy while reading questions, and best for privacy after interview, and overall preference). Only for ease of changing answers

<sup>10</sup> Reading level was not significant, except as noted in the footnotes of the tables. All significance tests were duplicated using Friedman's test, a nonparametric rank test which ignores the between-subjects factors. There were no appreciable changes in any of the  $p$ -values using this test.

Table 2. Mean subjective ratings and objective indicators of performance by mode of administration

	Audio <sup>5</sup> CASI	Video CASI	Paper SAQ	<i>p</i>
<i>Subjective ratings (0–2 scale)</i>				
Liked best	1.53	1.23	0.24	< 0.01 <sup>1</sup>
Best for asking sensitive Qs.	1.29	1.21	0.50	< 0.01 <sup>1</sup>
Easiest to change answer	0.96	0.93	1.14	0.61
Most interesting	1.72	1.11	0.15	0.01 <sup>2</sup>
Easiest to use	1.57	0.91	0.53	< 0.01 <sup>2</sup>
Best for getting honest answers	1.30	1.16	0.54	< 0.01 <sup>1</sup>
Best for privacy after interview	1.37	1.36	0.27	< 0.01 <sup>1</sup>
Best for privacy during interview	1.34	1.07	0.60	< 0.01 <sup>1</sup>
Overall preference	1.38	1.10	0.51	< 0.01 <sup>2</sup>
<i>Objective indicators</i>				
Minutes to complete	9.26	9.03	8.88	0.91
Requests for help	0.67	1.21	0.44	0.02 <sup>3</sup>
Proportion of correct skips	1.00	1.00	0.79	< 0.01 <sup>1</sup>
Ns <sup>4</sup>	35	35	35	

Notes. The first eight rows of this table report means of respondents' rankings, given on a scale of 0 (low) to 2 (high).

<sup>1</sup>Mean for Paper SAQ was different from means for Video-CASI and Audio-CASI at  $p < 0.05$  by *t*-test. Means for Video-CASI and Audio-CASI were not significantly different from each other by the same test. There was no significant association between reading level and mean score.

<sup>2</sup>Mean for Paper SAQ was different from means for Video-CASI and Audio-CASI at  $p < 0.05$  by *t*-test. Means for Video-CASI and Audio-CASI were different from each other at  $p < 0.05$  by *t*-test. There was no significant association between reading level and mean score.

<sup>3</sup>Mean for Paper SAQ was different from mean for Video-CASI (but not Audio-CASI) with  $p < 0.05$  by *t*-test. A significant association was found between reading levels and scores ( $p < 0.05$ ). (There were, on average, more requests for help from below average readers [1.02] than from average readers [0.51].)

<sup>4</sup>Ns shown are the *minimum* sample sizes for calculation of any proportion shown in the column.

<sup>5</sup>In the Audio-CASI administration, questions were displayed on the PC screen simultaneously with their audio presentation. Respondents had the option of turning off this video display, but none did so.

was the paper SAQ judged equivalent to the CASI modes. Between the two CASI modes, Audio-CASI was rated superior ( $p < 0.05$ ) for overall preference, ease of use, and most interesting to use. Other differences between the two modes in these subjective ratings were not statistically significant although they did consistently favor Audio-CASI.

#### 4.3.2. Reporting of drug use

Our analysis of the reporting of drug use focused on two issues: data quality and levels of sensitive behaviors reported. With regard to data quality, the computerized questioning produced higher quality data. Across all drug questions, subjects using the computer (both with and without the voice feature) gave only one "don't know"

Table 3. Proportion of respondents reporting use of cigarettes, alcohol and illicit drugs by mode of administration

	Audio CASI	Video CASI	Paper SAQ	<i>p</i>
<b>Cigarettes</b>				
Past 30 days	0.43	0.58	0.46	0.82
Past 12 months	0.43	0.67	0.46	0.65
Ever in lifetime	0.93	1.00	0.69	0.02 <sup>1</sup>
<b>Alcohol</b>				
Past 30 days	0.43	0.58	0.46	0.82
Past 12 months	0.64	0.75	0.62	0.63
Ever in lifetime	0.86	0.92	0.77	0.35
<b>Marijuana</b>				
Past 30 days	0.21	0.17	0.00	0.09 <sup>1</sup>
Past 12 months	0.29	0.50	0.08	0.04 <sup>1</sup>
Ever in lifetime	0.64	0.83	0.46	0.10 <sup>1</sup>
<b>Cocaine</b>				
Past 30 days	0.00	0.00	0.00	—
Past 12 months	0.07	0.08	0.00	0.31
Ever in lifetime	0.29	0.33	0.00	0.03 <sup>1</sup>
<b>Crack</b>				
Past 30 days	0.00	0.00	0.00	—
Past 12 months	0.07	0.17	0.00	0.20
Ever in lifetime	0.21	0.17	0.00	0.09 <sup>1</sup>
Ns <sup>2</sup>	14	12	13	

Notes. In Audio-CASI administration, questions were displayed on PC screen simultaneously with their audio presentation. (Respondents had option of turning off this video display, but none did so.)

<sup>1</sup>Paper SAQ different from Video-CASI and Audio-CASI at  $p < 0.10$  by *t*-test; Video-CASI and Audio-CASI not significantly different from each other by same test.

<sup>2</sup>Ns shown are the *minimum* sample sizes for calculation of any proportion shown in the column.

answer (one subject did not know how many days she had used marijuana in her lifetime). On the other hand, subjects answering paper and pencil drug questions left a total of 28 answers blank that they should have answered, and answered 22 additional questions unnecessarily due to failure to follow "skip" instructions.

Reporting of sensitive behaviors was examined in an exploratory analysis. Due to the small sample sizes (especially for

questions where only experienced subjects answer the questions), we did not expect that statistically significant findings at the  $p < 0.05$  level would be found. For exploratory purposes, we adopted the level of  $p < 0.10$  level as our criterion of significance. This significance level is indicative of a trend that merits examination in future studies.

The results for the reporting of drug use can be found in Table 3. Our analyses

focused on simply counting the numbers of subjects who reported any use of these drugs in the three reference periods. Questions about tobacco and alcohol consumption generally showed no differences among the three modes of questioning, as seen in Table 3. The only significant finding was that more subjects reported ever having smoked cigarettes in the computerized formats (100% and 93% for Video- and Audio-CASI) than on the paper SAQ (69%). For illegal drugs (marijuana, cocaine, and crack), the reported prevalence of drug use was much lower, but there was a trend for more subjects to report illegal drug use in the computerized formats (Video- and Audio-CASI) than in the paper SAQ format. No significant differences were observed between the Video- and Audio-CASI formats.

For marijuana use, none of the 13 subjects answering drug questions on paper SAQ reported using marijuana within the past 30 days, while a total of 5 did so in the computer assisted formats (3 of 14 for Audio-CASI and 2 of 12 for Video-CASI). This result is significant using a  $2 \times 2 \chi^2$  analysis (CASI formats vs. paper SAQ) at  $p = 0.090$ . One of 13 subjects reported using marijuana in the past year on the paper SAQ, while 10 of 26 so reported for the computerized questioning. This result was significant using a  $2 \times 2 \chi^2$  analysis (CASI formats vs. paper SAQ) at  $p = 0.044$ . Six of the 13 subjects reported lifetime marijuana use with the paper SAQ, while 19 of 26 did so in the computerized formats ( $p = 0.098$ ).

For cocaine and crack use, none of the subjects who received the paper SAQ reported ever using these drugs in their lifetimes. For the computerized formats, none of the 26 subjects reported using either drug in the past 30 days. However, 3 of 26 reported using crack within the past year

and 2 of 26 reported using cocaine within the past year (neither of these levels is significantly different from the 0.0% reporting use on the paper SAQ). For lifetime use, 8 of 26 subjects reported cocaine use, and 5 of 26 reported lifetime crack use ( $p$ -values from the same type of  $2 \times 2 \chi^2$  analysis,  $p = 0.025$  and  $p = 0.090$ , respectively).

#### 4.3.3. Reporting of sexual behavior

Data quality for the sex questions was generally good. Subjects answering computerized questions refused to answer a total of only three questions, and similarly, subjects using paper SAQs left only three questions unanswered. However, subjects using the paper SAQs mistakenly answered many inappropriate sex questions as a result of failure to follow instructions to skip over questions. Subjects answering the paper SAQs answered an average of 22.4 such inappropriate questions.<sup>11</sup>

Table 4 summarizes subjects' responses to the questions about sexual behavior obtained by the three modes of data collection.<sup>12</sup> Table 4 presents results for men's and women's heterosexual activities. It should be noted that men but not women were asked about sex with prostitutes. Furthermore, men were asked about paying for sex and forcing a woman to have sex against her will while the corresponding questions to women asked about being paid for sex and being forced to have sex against her will. Finally, only women were

<sup>11</sup> Across all three questions contents, there were non-significant differences between average and below average reading ability subjects in the mean numbers of extra questions that were answered (means of 3.3 versus 13.3). These differences were not statistically significant ( $t(38) = 1.54$ ), due to relatively high variance in the number of extra questions answered by subjects of low reading ability. (A couple of respondents answered all the questions.)

<sup>12</sup> Note that questions about activities with subjects' second and third most recent partners were not included in these analyses.

Table 4. Proportion of respondents reporting various sexual and fertility-related behaviors by mode of administration

	Audio <sup>5</sup> CASI	Video CASI	Paper SAQ	<i>p</i>
Heterosexual sex	1.00	1.00	0.92	0.13
Vaginal intercourse	1.00	1.00	1.00	–
Oral sex from partner	0.60	0.55	0.31	0.22
Oral sex to partner	0.71	0.73	0.46	0.17
Anal intercourse	0.07	0.25	0.15	0.96
Other heterosexual acts	0.20	0.50	0.15	0.23
Steady relationship now	0.73	0.75	0.54	0.20
Sex with IV drug user	0.13	0.25	0.15	0.81
Sex with bisexual	0.13	0.08	0.23	0.32
Sex with casual date	0.33	0.25	0.31	0.94
Forced to have sex	0.33	0.17	0.08	0.18
Paid for sex	0.27	0.08	0.23	0.74
Discuss STDs with most recent partner	0.80	0.58	0.33	0.03 <sup>1</sup>
Discuss birth control with most recent partner	0.80	0.75	0.50	0.08 <sup>1</sup>
Ns <sup>3</sup>	15	12	12	
<i>Women only</i>				
Ever pregnant	0.83	0.89	0.82	0.77
Given birth <sup>2</sup>	0.90	0.88	1.00	0.29
Stillbirth/miscarriage <sup>3</sup>	0.20	0.25	0.11	0.48
Abortion <sup>3</sup>	0.30	0.25	0.11	0.33
Ns <sup>3</sup>	10	8	9	

Note. Proportions for reporting of male–male sexual contacts and contacts with prostitutes (asked of men only) are not shown because only 5 men were included in the study sample. Note also that in Audio-CASI administration, questions were displayed on PC screen simultaneously with their audio presentation. (Respondents had option of turning off this video display, but none did so.)

<sup>1</sup>Paper SAQ different from CASI and Audio-CASI at  $p < 0.10$  by *t*-test; CASI and Audio-CASI not significantly different by same test.

<sup>2</sup>Questions asked of women who reported being pregnant.

<sup>3</sup>Audio-CASI and Video-CASI administrations were programmed to ask about abortions and miscarriages, etc. *only* when the number of pregnancies exceeded the number of children. In the paper and pencil format, these questions were asked of all women. Thus the data from these two formats is not directly comparable. For exploratory purposes in this analysis, we imputed a value of “no” for abortions and miscarriages for all instances in which the CASI formats did not ask about the number of abortions or miscarriages, etc.

<sup>4</sup>Ns shown are the *minimum* sample sizes for calculation of any proportion shown in the column.

<sup>5</sup>In the Audio-CASI administration, questions were displayed on the PC screen simultaneously with their audio presentation. Respondents had the option of turning off this video display, but none did so.

asked about pregnancy outcomes. Table 4 presents results for questions asked in a yes-no format and reports  $\chi^2$  analyses of differences in reporting of behaviors across groups. (Analyses of results for quantitative questions asking about the frequency of particular sexual activities showed no systematic differences among the different modes of questioning.)

The answers to yes-no questions about heterosexual activities did not vary systematically as a function of mode of questioning, as seen in Table 4. Only two questions (whether birth control and sexually transmitted diseases were discussed with most recent partner) yielded significant differences in response by mode of administration,  $p = 0.030$  and  $p = 0.083$ , respectively, in the  $2 \times 2 \chi^2$  analysis. Weak suggestions of a mode difference are also found for the questions in active oral sex ( $p = 0.17$ ) and forced sex ( $p = 0.18$ ). In both instances, the CASI formats obtained higher reported prevalence rates.

#### 4.3.4. Reporting of income

Data quality indicators for the income questions suggest that the computerized modes of data collection collected better quality data. The 24 subjects answering computerized versions of the questions entered 9 "don't know" answers, 1 refusal, and 3 "not sure" answers. For the income questions, "not sure" answers were followed in the CASI formats by additional prompting designed to help subjects understand the question. For 2 of the 3 "not sure" answers, these prompts appeared to facilitate subjects' decisions that they had, indeed, received the forms of income asked about in the questions. For the 14 subjects completing paper SAQs there was considerably more missing data: a total of 32 blank or uninterpretable answers were given by

these 14 subjects. These subjects also answered 4 questions they should have skipped.

Substantive responses to the income questions did not differ significantly as a function of mode of questioning. Only a single question showed any significant differences in levels of behavior under the different modes of questioning. The numbers of subjects who actually received each type of income was small, due to relatively young ages of the subjects.

#### 4.3.5. Summary of statistical findings

On almost all subjective indicators, subjects preferred CASI administration to paper SAQs. The sole exception to this rule was for ease of changing answers; paper SAQs were rated superior (nonsignificantly) on this attribute. Subjects' subjective ratings of Audio-CASI were consistently higher than those of Video-CASI; however, this result was significant for only three rating categories: overall preference, interest, and ease of use. There were neither substantial nor significant differences in the length of time required to complete interviews using the three methods, and, as expected, subjects using the paper SAQ did make a substantial number of errors in following "skip" instructions (21% incorrect). These instructions were, of course, flawlessly executed in CASI. Unexpectedly, Video-CASI engendered significantly more requests for assistance than either Audio-CASI or paper SAQs.

Data quality for the two computerized formats was superior to that of the paper SAQs. Subjects left more questions unanswered and they answered unnecessary questions more often with the paper SAQs. Because the computer skipped over unnecessary questions, it relieved subjects of the responsibility for correctly interpreting branching instructions. Furthermore,

because CASI subjects had to indicate that they did not know the answer, or that they refused or were unsure before going on to the next question, they produced clearer indications of why they were not giving useful answers. With the paper SAQ, subjects left many answers blank, requiring the researcher to guess why. Since the confidential nature of the questioning precluded reconciling such problems with the subject, help from the subjects themselves was not forthcoming.

The frequency of sensitive behaviors reported by subjects was not uniformly affected by the mode of data collection. However, when differences were observed, they generally showed higher levels of disclosure with the computer assisted self-interviewing formats. Thus subjects answering computerized versions of the questions reported significantly higher frequencies of marijuana use within the past month ( $p < 0.10$ ), within the past year, and over their lifetimes. They also reported higher levels of lifetime cocaine and crack use (subjects answering paper SAQs reported no use of these drugs). For sexual behavior questions, there were only two results that exceeded chance expectations at the  $p < 0.10$  level. Men and women were more likely to report having discussed birth control and protection from sexually transmitted diseases with their most recent sexual partners when asked about these behaviors by computer. There were also weak trends ( $p < 0.20$ ) for subjects to be more likely to report active oral sex and forced sex using the CASI formats.

Taken together, these results suggest that the use of self-administered computerized questioning offers the possibility of getting better quality data and possibly higher levels of disclosure of sensitive behaviors. Perhaps the perceived superior privacy of

the computer over paper SAQs or the greater ease of using the computer are contributing factors. Further research is currently underway to replicate and extend these preliminary findings; these efforts include:

- Methodological pretesting of an Audio-CASI component for the next round of the National Survey of Family Growth which will be conducted in 1994 by RTI and NCHS (O'Reilly 1993; Weeks 1993);
- A four-year program of basic research on the use of Audio-CASI technology for the survey measurement of adult sexual and contraceptive behaviors funded by the National Institutes of Health (Turner, Forsyth, Biemer, and La Vange 1993);
- Methodological testing of the use of Audio-CASI to assess risk and sexual behaviors of a national sample of adolescents.

## 5. Conclusion

The performance and results obtained using our Audio- and Video-CASI systems were quite encouraging. The system hardware and software operated satisfactorily, and a range of less-educated and computer-inexperienced subjects successfully operated the system after only a brief introduction. Overall, subjects preferred the two computer assisted formats (Audio- and Video-CASI), and the available evidence suggests that these formats not only eliminated errors due to faulty execution of skip instructions, but they also appeared to encourage more complete reporting of some sensitive behaviors. In the two instances where differences were observed between the two CASI formats, Audio-CASI was preferred by respondents and it

generated fewer requests for assistance than Video-CASI.

Our results suggest two general conclusions. First, CASI formats (both Audio- and Video-) can provide important advantages over paper SAQs. In addition to the advantages noted above, CASI formats eliminate the possibility of subjects becoming confused by complex "skip" logic that is built into a questionnaire. Second, in those instances where respondents' literacy is a problem, Audio-CASI offers a workable alternative mode of data collection that provides both a completely private mode of administration as well as the other advantages of computerized self-administration. Given the exploratory nature and very small sample sizes used in this present study, we would suggest that these conclusions be treated as the starting point for further methodological investigations.

## 6. References

- Ahsanullah, M. and Eichorn, B. (1984). On Scrambled Randomized Response of Sensitive Quantitative Data. Proceedings of the Survey Research Methods Section, American Statistical Association, 800-802.
- Baker, R.P. and Bradburn, N.M. (1991). CAPI: Impacts on Data Quality and Survey Costs. Paper presented at the 1991 Public Health Service Conference on Records and Statistics, Washington, DC, July 15-17.
- Bradburn, N.M. (1983). Response Effects. In Handbook of Survey Research, eds. P. Rossi, J. Wright, and A. Anderson, New York: Academic Press.
- Bradburn, N.M. and Sudman, S. (1979). Improving Interview Method and Questionnaire Design. San Francisco: Jossey-Bass.
- Bradburn, N.M., Frankel, M., Hunt, E., Ingels, J., Schoua-Glusberg, A., and Wojcik, M. (1993). A Comparison of Computer-Assisted Personal Interviews (CAPI) with Personal Interviews in the National Longitudinal Survey of Labor Market Behavior - Youth Cohort. Paper No. 7 in Information Technology in Survey Research Discussion Paper Series, Chicago: National Opinion Research Center.
- Camburn, D., Cynamon, M., and Harel, Y. (1991). The Use of Audio Tapes in Written Questionnaires to Ask Sensitive Questions during Household Interviews. Paper presented at the National Field Technologies Conference, San Diego, CA, May.
- Droitcour, J., Caspar, R., Hubbard, M., Parsley, T., Visscher, W., and Ezzatti, T. (1991). The Item Count Technique as a Method of Indirect Questioning: A Review of Its Development and a Case Study Application. In Measurement Errors in Surveys, eds. P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman, New York: John Wiley and Sons, Inc.
- Evan, W.M. and Miller, J.R. (1969). Differential Effects of Computer vs. Conventional Administration of a Social Science Questionnaire: An Exploratory Methodological Experiment. Behavioral Science, 14, 216-227.
- Groves, R.M. and Nicholls, W.L. (1986). The Status of Computer-Assisted Telephone Interviewing: Part II - Data Quality Issues. Journal of Official Statistics, 2, 117-134.
- Hay, D.A. (1990). Does the Method Matter on Sensitive Survey Topics. Survey Methodology, 16, 131-136.
- Johnston, G. (1992). Demonstration of Computer-Assisted Audio Survey Technology. Seminar presented at the

- National Center for Health Statistics, Hyattsville, MD, January 28.
- Jones, E.F. and Forrest, J.D. (1992). Under-reporting of Abortion in Surveys of U.S. Women: 1976 to 1988. *Demography*, 29, 113-126.
- Lessler, J. and Holt, M. (1987). Using Response Protocols to Identify Problems in the U.S. Census Long Form. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 262-266.
- Miller, H.G., Turner, C.F., and Moses, L.E., eds. (1990). Chapter 6: Methodological Issues in AIDS Surveys. In *AIDS: The Second Decade*. Washington DC: National Academy Press.
- National Center for Education Statistics (1993). *Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey*. Washington, DC: U.S. Department of Education.
- O'Reilly, J.M. (1993). *Lessons Learned Programming a Large, Complex, CAPI Instrument*. Presentation to Blaise Users Conference, October, 1993, London.
- O'Reilly, J.M., Lessler, J.T., and Turner, C.F. (1992a). Needs for the Demonstration of Audio Computer-Assisted Self Interviewing. Presentation to NIH and ADAMHA staff, National Institute of Child Health and Human Development, Rockville, MD, January 30, 1992.
- O'Reilly, J.M., Lessler, J.T., and Turner, C.F. (1992b). Needs for and Demonstration of Audio Computer-assisted Self Interviewing. Presentation to staff, U.S. Bureau of the Census, Suitland, MD, January 30, 1992.
- O'Reilly, J.M. and Turner, C.F. (1992). *Survey Interviewing Using Audio-Format, Computer-Assisted Technologies*. Presentation to the Washington Statistical Society, Washington, DC, March 18, 1992.
- Schwarz, N., Strack, F., Hippler, H., and Bishop, G. (1991). The Impact of Administration Mode on Response Effects in Survey Measurement. *Applied Cognitive Psychology*, 5, 193-212.
- Statistics Netherlands (1989). *Blaise 2.0 CAPI/CATI User Manual*. Voorburg, The Netherlands: Netherlands Central Bureau of Statistics.
- Turner, C.F., Forsyth, B.F., Biemer, P.P., and LaVange, L.M. (1993). *Survey Measurement of Sensitive Behaviors Using Audio-CASI: Evaluation of a New Technology*. Unpublished proposal for funded NIH Grant 1-R01-HD31067-001. Research Triangle Park, NC: Research Triangle Institute.
- Turner, C.F., Lessler, J.T., and Devore, J.W. (1992). Effects of Mode of Administration and Wording on Reporting of Drug Use. In *Survey Measurement of Drug Use: Methodological Studies*, eds. C.F. Turner, J.T. Lessler, and J.G. Gfroerer, DHHS Publication (ADM) 92-1929, Washington, DC: Government Printing Office.
- Umesh, U.N. and Peterson, R.A. (1991). A Critical Evaluation of Randomized Response Method Applications, Validation, and Research Agenda. *Sociological Methods and Research*, 20, 104-138.
- Weeks, M.F. (1993). *The National Survey of Family Growth: Start-up of a Complex CAPI Project*. Presentation at Southern Association for Public Opinion Research, Raleigh, NC, October.
- Weeks, M.F. (1992). *Computer-Assisted Survey Information Collection: A Review of CASIC Methods and Their Implications for Survey Operations*. *Journal of Official Statistics*, 8, 445-465.

Received July 1992  
Revised January 1994