IMPACT OF T-ACASI ON SURVEY MEASUREMENTS OF SUBJECTIVE PHENOMENA

THOMAS HARMON
CHARLES F. TURNER
SUSAN M. ROGERS
ELIZABETH EGGLESTON
ANTHONY M. ROMAN
MARIA A. VILLARROEL
JAMES R. CHROMY
LAXMINARAYANA GANAPATHI
SHEPING LI

Abstract   Numerous studies have shown that audio-computer-assisted self-interviewing (audio-CASI) and telephone audio-CASI (T-ACASI) technologies yield increased reporting of sensitive and stigmatized objective phenomena such as sexual and drug use behaviors. Little attention has been given, however, to the impact of these technologies on the measurement of subjective phenomena (attitudes, opinions, feelings, etc.). This article reports results for the seven subjective measurements included in

THOMAS HARMON was a student in the MA program in Applied Social Research at Queens College of the City University of New York (VUNY). He is presently a Program Coordinator at the League Treatment Center, 885 Rogers Ave., Brooklyn, NY 11226, USA. CHARLES F. TURNER is a Professor of Applied Social Research at CUNY (Queens College and the Graduate Center) and senior consultant to the Program in Health and Behavior Measurement at Research Triangle Institute (RTI), 2816 Northampton St., NW, Washington DC 20015, USA. SUSAN M. ROGERS AND ELIZABETH EGGLESTON are scientists with RTI's Program in Health and Behavior Measurement, 701 13th Street, NW, Washington DC 20005, USA. ANTHONY M. ROMAN is a Senior Fellow with the Center for Survey Research at the University of Massachusetts, Boston, 100 Morrissey Blvd, Boston, MA 02125, USA. MARIA A. VILLARROEL is currently a doctoral student in Epidemiology at Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205, USA. JAMES R. CHROMY, LAXMINARAYANA GANAPATHI AND SHEPING LI are with RTI, 3040 Cornwallis Road, RTP, NC 27709, USA. This article is based in part on an essay prepared by the first author for the M.A. degree in Applied Social Research at Queens College (CUNY). This work was supported by the National Institutes of Health [R01-MH56318 and R01-HD31067 to C. F. T.] The authors wish to thank Joseph A. Catania for collaboration in early phases of the design of this study and for access to the public use dataset from his 1996 National Survey of Sexual Health. The authors also wish to thank the many other people who made major contributions to this work during the proposal and design phase, most importantly, James Gribble. Address correspondence to Charles F. Turner; e-mail: CFTurner@PopEpi.org.

the National STD and Behavior Measurement Experiment (NSBME). NSBME drew probability samples of USA and Baltimore adults (Ns = 1,543 and 744, respectively) and randomized these respondents to be interviewed by T-ACASI or telephone interviewer-administered questioning (T-IAQ). Response distributions for all subjective measurements obtained by T-ACASI diverge from those obtained by human telephone interviewers. For six of our seven ordinal-scaled measurements, this divergence involved shifting responses directionally along the ordinal scale, as opposed to a nondirectional redistribution among response categories. When interviewed by T-ACASI, respondents were more supportive of traditional gender roles and corporal punishment, less supportive of integrated neighborhoods and same-gender sex, and more likely to agree that occasional marijuana use is harmless and to describe themselves as attractive. The majority of these results suggest that telephone survey respondents may provide more "tolerant" and "socially liberal" responses to human interviewers than to a T-ACASI computer. Similarly, although the evidence is not entirely consistent, the impact of T-ACASI appears to increase with the social vulnerability of the population surveyed.

Introduction

In the past two decades, computer-assisted self-interviewing (CASI) systems have come to dominate clinical and survey measurements of sensitive but objective measurements, such as AIDS-related sexual and drug-use behaviors. CASI systems fall into three main groups: video-CASI, audio-CASI, and telephone audio-CASI (T-ACASI). Video-CASI provided the earliest demonstration of the potential impact of automated self-interviewing on reducing underreporting of sensitive or stigmatized behaviors (Duffy and Waterton 1984). In a video-CASI application, respondents view questions on a computer screen and typically enter answers using the computer keyboard, mouse, or touchscreen. Audio-CASI adds audio features to those provided by video-CASI. In an audio-CASI interview, respondents listen to questions privately through headphones and enter responses by keyboard, mouse, or touchscreen. Early researchers and application developers saw this technology as a way of providing complete privacy to all respondents (including those with poor reading skills) when reporting on sensitive topics. T-ACASI adds telephone capabilities to a standard audio-CASI system thereby allowing audio-CASI interviews to be conducted over the telephone with the telephone's touchtone keypad serving as the input device.

A major motivation for the development of audio-CASI and T-ACASI technologies during the 1990s was the desire to reduce bias in survey and clinical measurements of the behaviors that spread HIV infection (Miller, Turner, and Moses 1990, Chapter 6; Turner and Sheon 1994; Turner, Lessler, and Gfroerer 1992). The measurements initially targeted by audio-CASI and T-ACASI tech-

nologies were hidden but objective phenomena that spread HIV infection, such as unprotected same-gender sexual contact among men, unprotected sex among persons of discordant or unknown infection status, injection drug use, and needle sharing (Fay et al. 1989; O'Reilly et al. 1994; Tourangeau and Smith 1996; Turner et al. 1996a, 1996b, 1998a, 1998b; DesJarlais et al. 1999; Anderson et al. 2006). Subsequent research has explored the impact of these technologies on measurements of a variety of other health behaviors (e.g., Epstein 2002; Brener et al. 2006). Although the evidence is not completely consistent, the weight of the available evidence suggests that in-person CASI methodologies probably do provide more accurate measurements of these sensitive behaviors.

The evidence is sparser for T-ACASI measurements of such behaviors. Our research group and collaborators have previously presented evidence that T-ACASI increased reporting of illicit (but not licit) drug use by both representative gay, urban male and general population samples; STD infection and both same-gender and heterosexual sexual behaviors in general population samples; and teen smoking in a regional sample (Gribble et al. 2000; Turner et al. 1996, 2005; Currivan et al. 2004; Villarroel et al. 2006a, 2008). The work of other researchers has yielded similar conclusions (Corkrey and Parkinson 2002; Lau, Tsui, and Wang 2003; Moskowitz 2004), although generalizations from some studies are restricted by methodological limitations, such as small sample sizes, use of specialized or volunteer populations, or very low response rates (e.g., Searles et al. 1995; Kobak et al. 1997; Millard and Carver 1999; Tourangeau et al. 2002).

The impact of T-ACASI on measurements of subjective phenomena is less certain because the required large-scale randomized survey experiments have not been conducted. This article reports the result of such an experiment.

RESEARCH STRATEGY

The primary research focus of the National STD and Behavior Measurement Experiment (NSBME) was "the assessment of: the extent, if any, to which new T-ACASI technology may increase the validity and reliability of telephone survey measurements of AIDS-related and other sensitive behaviors . . ." (Turner et al. 1995, p. 46). In the domain of such objective behaviors, the NSBME research team had many firm hypotheses about the impact of T-ACASI versus interviewer questioning. These hypotheses were derived from extensive prior research comparing results obtained by interviewer questioning versus more private interview modes (CASI and paper, self-administered questionnaires; see, for example, Turner, Lessler, and Devore 1992; Tourangeau and Smith 1996). In particular, we anticipated that T-ACASI would obtain more frequent reporting of same-gender sexual contacts, drug use, and incidence of sexually transmitted diseases, and less frequent reporting of protective behaviors such as consistent condom use—as found in prior research.

Fielding of the NSBME also provided an opportunity to expand the body of evidence on the impact of this technology on the neglected subjective domain. Toward this end, the NSBME research team added seven questions on subjective phenomena without making explicit predictions about the expected direction and magnitude of the mode effects. We thought that it was reasonable to expect that the more private interview mode afforded by T-ACASI would provide different and possibly more revealing insights into the subjective state of respondents.[1] We reasoned by analogy to our experience with sensitive sexual and drug-using behaviors. Just as some respondents would underreport injection drug use or same-gender sexual contacts when questioned by a human interviewer, so too, the social attitudes expressed by respondents might be influenced by the social context of the interview. We thus were prepared to find differences in a range of subjective measurements depending upon whether a human or computer was conducting the interview. As we will discuss after reporting our results, the *interpretation* of the "meaning" of these differences is neither straightforward nor amenable to definitive ascertainment.

While one might dismiss this research as "merely exploratory," we believe that would be a scientific mistake for two reasons. First, the results of this experiment provide the first assessments of the impact of this new technology on subjective measurements using a large, national, randomized experiment. Second, a failure to reject the null hypothesis of no difference between T-ACASI and telephone interviewer-administered questioning (T-IAQ) subjective measurements would be valuable given the dearth of reliable information on this increasingly common measurement technology. The NSBME sample size provides good statistical power to document even modest effects.

The NSBME surveyed probability samples of U.S. and Baltimore adults aged 18–45 and randomly assigned respondents to answer a wide range of questions asked either by a human telephone interviewer (T-IAQ) or by T-ACASI. In this article, we report NSBME results on the impact of survey mode on responses to the seven questions on subjective phenomena that were included in the NSBME. These questions cover a wide range of social and personal issues including the acceptability of corporal punishment, traditional gender roles in families, marijuana use, same-gender sex, residential segregation, and respondents' evaluation of their own attractiveness.

In this article, we seek answers to two questions. First, does T-ACASI technology obtain different distributions of responses to subjective measurements than those obtained by human telephone interviewers? Second, is the impact, if any, of T-ACASI technology on subjective measurements homogeneous across subpopulations defined by social and demographic factors?

---

1. This has sometimes been claimed for comparisons of interviewer-administered subjective measurements and those obtained by paper self-administered questionnaires (paper SAQs). As Sudman observed 40 years ago: "Where there are large differences [between results from paper SAQs and personal interviews] the self-administered forms seem to give a better measure of the true feelings of respondents than do the personal interviews . . ." (Sudman et al. 1965, p. 297, emphasis added).

Methods[2]

SAMPLE AND EXPERIMENTAL DESIGN

As described more completely in prior publications,[3] the NSBME was embedded in a telephone survey of a probability sample of women and men aged 18–45 years residing in U.S. households with working landline telephones. The survey was conducted between September 1999 and April 2000. Two sample strata were recruited: (1) a sample of the telephone-accessible U.S. household population aged 18–45 (national stratum) and (2) a parallel sample of the telephone-accessible population of the city of Baltimore, MD (Baltimore stratum). A list-assisted random digit dialed (RDD) sample was drawn for each stratum using the Genesys Sampling System (2002).

For the national stratum, 14,250 telephone numbers were generated and 12,322 (86.5 percent) were successfully screened for eligibility. Of these screened telephone numbers 2,183 were found to be residential numbers with one or more eligible English-speaking respondents aged 18–45. One eligible household member of these households was randomly selected for participation in the survey (without substitution). Of the 2,183 target respondents in the national stratum, 1,452 (66.5 percent) completed interviews, and 91 respondents (4.2 percent) completed partial interviews that included at least one substantive questionnaire section.

The second sample stratum was drawn to represent the adult population of Baltimore, MD to permit comparisons with an earlier study (Turner et al. 2002). For this stratum, 7,498 telephone numbers were generated and 6,326 (84.4 percent) were successfully screened for eligibility. Screening identified 1,072 households with an eligible respondent, and 697 of these eligible respondents completed interviews (65.0 percent). An additional 47 respondents (4.4 percent) completed partial interviews.

AAPOR (2006) response rates (formula RR3) for the NSBME were 62 percent for the T-IAQ condition and 53 percent for the T-ACASI condition in the national stratum. In the Baltimore stratum, these response rates were 56 percent for the T-IAQ condition and 50 percent for the T-ACASI condition (see Roman 2000, pp. 23–29).

INTERVIEW MODES

Telephone numbers were randomly assigned to the T-IAQ or T-ACASI conditions prior to their release to the telephone survey unit. Following screen-

---

2. The protocol for this research was approved and supervised by Institutional Review Boards at the Research Triangle Institute (RTI) and the University of Massachusetts at Boston.
3. This description of NSBME methodology is excerpted from longer descriptions published elsewhere (Roman 2000; Villarroel et al. 2006a, 2006b, 2008).

ing and recruitment into the study, telephone interviewers at the Center for Survey Research (University of Massachusetts, Boston) conducted the survey either by asking the respondent questions and recording their answers (T-IAQ condition) or by transferring the respondent to the T-ACASI system (Cooley et al. 2000). Both male and female interviewers were used in both conditions.

SURVEY MEASUREMENTS

The NSBME included seven measurements of unquestionably subjective phenomena. (The Appendix presents complete question wordings.) Five of these items were adapted from the General Social Survey (GSS; Davis, Smith, and Marsden 2002). These items asked about approval of corporal punishment for children; approval of same-gender sexual relations; preference for residing in racially segregated or racially integrated neighborhoods; and two questions asked about approval of "traditional" gender roles in families (i.e., wives should take care of home and family while husband should be achiever outside the home; and preschool children suffer if mothers work). The NSBME included a sixth question asking respondents to assess how "good looking" they were compared to the average man or woman; this question was adapted from the National Survey of Adolescent Males (Sonenstein et al. 1998). The seventh question asked respondents whether they agreed or disagreed that occasional use of marijuana posed "no real risk"; this item was proposed by a member of the research team (SR) for an earlier survey (Turner et al. 2002). These seven questions[4] were asked near the end of the NSBME interview after most respondents had answered approximately 30 minutes of questions focused mainly upon sexual behaviors, drug use, and STD history.

STATISTICAL ANALYSIS

Our analyses of the NSBME are intended to answer two related questions. First, does the mode of interview affect the responses given to these seven questions on subjective phenomena? Second, is the impact, if any, of the interview mode on reporting of these subjective phenomena homogeneous across subpopulations (defined by gender, ethnicity, etc.) or are some subpop-

4. These subjective measurements did not appear in sequence. The first question about personal attractiveness was numbered 75, and it followed 11 nonsexual questions asking about experiences with domestic violence, corporal punishment, arrests, unemployment, and receipt of welfare. The next questionnaire section was prefaced by the statement "Now we have some questions about how you see yourself and about your feelings and moods." After the first question on attractiveness, three questions were asked about respondents' symptoms, medication, and treatment for anxiety or depression. The six remaining subjective measurements followed these items. Respondents in the T-ACASI mode who answered the entire questionnaire took an average of 29 minutes (SD = 4.3 minutes).

ulations particularly sensitive to the interview mode? To address our research questions, we combine the national and Baltimore sample strata. The combined sample strata are treated as a population that has been randomly allocated to one of two experimental conditions: T-ACASI or T-IAQ interview mode. Data in these analyses are unweighted, and our statistical analyses assess the likelihood that observed fluctuations in survey responses across the two interview modes arose by chance from the random allocation of respondents to one of the two experimental groups.

Likelihood ratio chi-square tests were performed to test the equivalence of the response distributions obtained in the T-ACASI and T-IAQ conditions, and Mantel–Haenszel tests for linear association were performed to test the impact of T-ACASI when the categories of response (e.g., strongly agree, agree, disagree, strongly disagree) were treated as an ordinal scale. These analyses were carried out using SPSS version 6.0.1 (SPSS Inc., 1993) and Stata SE, version 8 (Stata Corporation, 2003). To provide an explicit control for any impact of variation in sample composition between the T-ACASI and T-IAQ conditions due to differential sample attrition or other factors, we also fit ordered logit models (Long and Freese 2001, Chapter 5) to test for T-ACASI mode effects controlling for respondent gender, sample strata, race (Black or non-Black), Hispanic origin, age and education in years, marital status (married or living together versus other), region of country (six regions), and urbanicity (four categories of the population density). (See footnotes to table 1 for additional details.)

Ordered logit models and hierarchical log-linear models (Goodman 1968) were subsequently used to test whether the impact of T-ACASI was equivalent across subpopulations defined by gender, race, region of residence, and other social and demographic factors. Ordered logit models were used to conduct these tests for subjective measurements for which we found a statistically significant or borderline ($p < .10$) estimated effect of T-ACASI when the subjective measurement was treated as an ordinal variable. These analyses used the maximum-likelihood ordered logit estimation procedures of Stata SE, version 8. For any variable that evidenced a statistically reliable or borderline T-ACASI mode effect ($p < .10$) when the subjective measurement was treated as a nominal level variable—but no such effect when the measurement was treated as an ordinal variable—we fit log-linear models to three-way tables of subjective measurement (*S*: four or five categories, e.g., strongly agree, agree, disagree, strongly disagree) *by* mode of interview (*M:* T-IAQ or T-ACASI) *by* sociodemographic factor *(F*, for example, male or female). This log-linear modeling was carried out using statistical software from SPSS, version 6.0.1.

Table 1. Responses to Seven Measurements of Subjective Phenomena by Mode of Interview: Analysis of Unweighted Data from the NSBME

Measurement and survey mode	Response[e]				N	p[d]			Item nonresponse[f]	
	Strongly agree	Agree	Disagree	Strongly disagree		Nominal	Ordinal	Ordinal adjusted[e]	Percentage	p
Sometimes necessary to give children a good, hard spanking										
T-ACASI	11.9%	34.1%	36.2%	17.8%	900				2.9	
T-IAQ	8.2%	39.4%	32.2%	20.2%	1,151	.002	.245	>.500	1.0	.002
There is no real risk in occasional pot use										
T-ACASI	10.5%	20.2%	38.5%	30.8%	877				5.5	
T-IAQ	6.9%	22.7%	35.6%	34.8%	1,139	.006	.043	.067	2.1	<.001
Preschool children suffer if mother works										
T-ACASI	13.2%	27.5%	48.3%	10.9%	906				2.3	
T-IAQ	9.9%	30.7%	45.4%	13.9%	1,149	.010	.089	.127	1.2	.06
Much better if man is achiever outside home and woman takes care of home and family										
T-ACASI	11.0%	20.6%	44.7%	23.7%	907				2.2	
T-IAQ	8.0%	22.0%	42.4%	27.6%	1,152	.030	.032	.037	1.0	.023
Opinion about sex between adults of the same gender	*Always wrong*	*Almost always wrong*	*Sometimes wrong*	*Not wrong at all*	N	*P nominal*	*P ordinal*	*Ordinal adjusted*[e]		
T-ACASI	49.1%	7.5%	12.7%	30.7%	872				5.9	
T-IAQ	46.0%	5.3%	11.0%	37.8%	1,120	.005	.012	.014	3.7	.016

Measurement and survey mode	Response[e]					p[d]			Item nonresponse[f]	
	Strongly agree	Agree	Disagree	Strongly disagree	N	Nominal	Ordinal	Ordinal adjusted[e]	Percentage	p
Compared to an average person, how good looking are you?[a]	*Much better*	*Somewhat better*	*Average*	*Worse looking*[a]	*N*	*P nominal*	*P ordinal*	*Ordinal adjusted*[e]		
T-ACASI	16.8%	38.3%	41.9%	3.0%	906				2.3	
T-IAQ	10.2%	31.8%	55.6%	2.3%	1,152	<.001	<.001	<.001	1.0	.015
Preferred racial composition of neighborhood, if you could live anywhere (Black and White respondents only[b])	*All same race*	*Mostly same race*	*50-50*	*All or mostly other race*	*N*	*P nominal*	*P ordinal*	*Ordinal adjusted*[e]		
T-ACASI	14.8%	46.7%	36.1%	2.3%	783				2.9	
T-IAQ	10.4%	37.6%	48.9%	3.1%	931	<.001	<.001	<.001	7.7	<.001

[a] Response categories of "somewhat worse than average" (2.4 percent of total sample) and "much worse than average" (0.2 percent) have been combined in this analysis.

[b] This analysis was restricted to respondents who identified their race as either White or Black to simplify interpretation. The question characterizes neighborhoods as all Black, mostly Black, half Black and half White, mostly White, or all White.

[c] Distribution of substantive responses excludes nonsubstantive responses, e.g., "don't know," refused, no answer, etc. See the final two columns for percents giving such nonsubstantive responses.

[d] *p*-values were calculated from bivariate likelihood ratio chi-square tests, bivariate Mantel–Haenszel tests for linear association, and multivariate ordered logit models including nine sociodemographic control variables.

[e] Ordered logit models were fit to each measurement. These models treated the response (e.g., strongly agree, agree, disagree, strongly disagree) as ordinal variables and estimated the impact of T-ACASI (versus T-IAQ) adjusting for impact of gender, sample strata (Baltimore versus USA), race (Black or non-Black), Hispanic origin, age in years, education in years, marital status (married or living together versus other), region of country (Northeast, South Atlantic, North Central, South Central, Mountain, and Pacific states), and urbanicity (21 largest MSAs; counties with populations of 85,000 or more; counties with populations of 20,000–84,999; and counties with populations of less than 20,000). (While sample strata were introduced in regressions, it was dropped during estimation due to apparent redundancy with the Region and Urbanicity indicators.) *p*-values shown test the null hypothesis that the estimated T-ACASI effect (net of impact of control variables) is zero.

[f] This table displays percentage of respondents who did not give a substantive answer to each question among respondents who did answer at least one of the three questions that preceded the seven subjective measurements. (This was done to eliminate interview breakoffs from the analysis. Prior analyses (Villarroel et al. 2006a) have reported that there was a higher rate of interview breakoff in the T-ACASI condition of the NSBME but this did not compromise the conclusion that T-ACASI significantly shifted the substantive response distributions.) *p*-values in this column are calculated with Pearson's chi-square tests.

Results

SAMPLE EQUIVALENCE ACROSS INTERVIEW MODES

Previously published analyses (Villarroel et al. 2006a, 2006b) have demonstrated that the demographic characteristics of respondents completing the T-IAQ and T-ACASI versions of the NSBME questionnaire were statistically equivalent. (Tests of the equivalence of T-IAQ and T-ACASI samples by gender, age, marital status, education, race/ethnicity, region, urbanization, and sample strata produced no statistical result with a *p*-value less than .28.)

ITEM NONRESPONSE

The rightmost columns of table 1 display the percent of respondents in each survey mode who gave nonsubstantive responses (e.g., "don't know," refuse, no answer) to the seven subjective measurements. For six of the seven questions, item nonresponse rates were one to three percentage points higher in the T-ACASI condition. While these effects are small, almost all of them are statistically reliable. A somewhat larger effect in the opposite direction is found for the question on subjects' preferred racial composition of their ideal neighborhood of residence. When questioned by a human interviewer, 7.7 percent of respondents did not provide a response, while only 2.9 percent failed to respond when questioned by a T-ACASI computer ($p < .001$). This mode effect did not vary for Black and White respondents.[5]

IMPACT OF T-ACASI ON SUBSTANTIVE RESPONSES: BIVARIATE ANALYSIS

Table 1 displays the distribution of substantive responses obtained for the seven subjective measurements by survey modes. For each measurement, we find a statistically reliable difference in the distribution of responses obtained by the two survey modes when the categories of the subjective measurement are treated as nominal variables. When the response categories are treated as ordinal measurements, the same general result[6] holds for all measurements except for endorsement of "good, hard spanking[s]" for young children. For this measurement, the lack of a significant ordinal effect is explained by the fact that there are statistically significant or borderline differences within the two categories of agreement ($p = .001$) and disagreement ($p = .053$),

---

5. There was a 4.8 percentage point difference in nonresponse rates across modes for White respondents and a 4.5 percentage point difference for Black respondents. For both groups, item nonresponse was higher in the T-IAQ condition.

6. The *p*-value for the Mantel–Haenszel test for the linear effect of T-ACASI on opinions of the impact of mother's working declines to borderline significance ($p = .089$). The chi-square test of independence for the two- by four-category table (TACASI by four categories of agreement) has a *p*-value of .010.

while the overall agree–disagree ratios are statistically equivalent (46 percent agreement in T-ACASI and 47.6 percent in T-IAQ; $p = .493$; see table 1).

For the other six measurements, the Mantel–Haenszel tests for linear association indicate that T-ACASI shifted responses to a higher (or lower) ordered category—although in one instance, this result only approaches statistical significance ($p = .089$). Overall, survey respondents interviewed by T-ACASI expressed greater approval of traditional gender roles (women should be homemakers, and preschool children suffer if their mother works), greater disapproval of same-gender sex, and preference for residential neighborhoods with more people of the same race as the respondent's. T-ACASI respondents also expressed greater agreement that there is no risk to occasional marijuana use, and they described themselves as better looking than respondents interviewed by human interviewers.

IMPACT OF T-ACASI: MULTIVARIATE CONTROLS

Randomization in this experiment provides a strong basis for the inference that the observed discrepancies in our subjective measurements are due to the mode of interview. Previous findings that the observed distributions of sociodemographic characteristics were approximately equivalent across experimental conditions in the NSBME are reassuring in this regard. Nonetheless, the nonequivalence in response rates in the two survey modes remains a source of concern. To address this issue, we fit multivariate ordered logit models in which the T-ACASI effects were estimated net of the impact of nine control variables: gender, sample strata (Baltimore versus USA), race, Hispanic origin, age, education, marital status, region of country, and urbanicity. As shown in table 1 (see the column labeled "Ordinal adjusted"), results of these tests were quite similar to those obtained in our bivariate analyses. So, for example, when treated as an ordinal measurement, responses to the spanking question do not exhibit a significant T-ACASI mode effect with or without our multivariate statistical controls. Similarly, statistical tests for the other six subjective measurements yielded remarkably similar p-values in the bivariate and multivariate tests—although p-values for responses to the marijuana question crossed the .05 boundary ($p = .043$ for bivariate analysis versus $p = .067$ for multivariate analysis).

SUBPOPULATION DIFFERENCES IN THE IMPACT OF T-ACASI: ORDERED LOGIT MODELS

We separately tested the homogeneity of the impact of T-ACASI across groups defined by gender, race, age, education, marital status, urban residence, region of the USA, and sample stratum (USA or Baltimore).[7] In addition, for

7. Note that this analysis estimated a series of models entering sociodemographic factors one at a time. This analysis answers the question of whether the magnitude of the estimated mode effect varies across groups defined by that factor. We believe that this is a more useful exploration of these results than an attempt to model the simultaneous impact of all factors.

judgments of the risk of occasional marijuana use, we tested the homogeneity of estimates of T-ACASI impact for respondents who reported ever using marijuana versus those who did not.

For the six subjective measurements for which the Mantel–Haenszel tests of linear effect were significant, we fit a series of ordered logit models that included a binary variable identifying the experimental condition (T-ACASI versus T-IAQ), the sociodemographic variable, and an interaction term to capture variation in the T-ACASI effect across levels of the sociodemographic variable. Age and education (in years) were entered as metric independent variables; the remaining sociodemographic factors were entered as one or more binary variables (see footnotes to table 2).

Four sociodemographic variables demonstrated statistically significant or borderline ($p \leq .10$) heterogeneity in the estimated impact of T-ACASI on two or more subjective measurements. Three of these variables show consistent patterns of heterogeneity in the impact of T-ACASI. First, the impact of T-ACASI weakens as the respondent's level of education increases for questions on the risk of marijuana use, preferred racial composition of the neighborhood of residence, and respondents' ratings of their own attractiveness. Second, the impact of T-ACASI weakens as the respondent's age increases for questions on the preferred racial composition of the neighborhood of residence, one of the two questions on traditional gender roles ("preschool child is likely to suffer if his or her mother works"), and respondents' ratings of their own attractiveness. Third, the impact of T-ACASI is stronger in the Baltimore sample stratum than in the national sample stratum for the preferred racial composition of respondents' ideal neighborhood, and for both questions on the endorsement of traditional gender roles ("it is much better . . . if the man is the achiever outside the home and the woman takes care of the home and family"; and for the "preschool child . . . suffers . . ." question).

There are also two instances of significant heterogeneity in estimates of the impact of T-ACASI by gender, but the results are inconsistent. T-ACASI has a stronger mode effect for men than women in assessments of approval of same-gender sex, but the reverse is true for measurements of respondents' preferred racial composition of their neighborhood of residence. There are three additional instances in which significant or borderline heterogeneity was found for estimates of T-ACASI mode effects for a single subjective measurement. For estimates of the risk of occasional marijuana use,[8] an insubstantial T-ACASI mode effect was found for *non-Blacks* and for persons who reported *never* smoking marijuana (gammas for ordinal association between survey mode and assessed risk were −0.01 and 0.02, respectively) while a more substantial effect was found for *Blacks* and for respondents reporting *prior marijuana use* (gammas = 0.20 and 0.11; positive gammas indicate that T-ACASI shifted

8. $p = 0.02$ for the test of three-way interaction with race and .08 for interaction with prior marijuana use.

responses toward *disagreement* that occasional marijuana use was risky). Finally, the estimated T-ACASI mode effect on the endorsement of the view that preschool children suffer if their mother works showed a significant variation ($p = .05$) by whether or not the respondent was married or cohabiting. A minuscule mode effect was found among persons who were married or cohabiting (gamma = 0.02) while a stronger effect in the opposite direction was found among other respondents (gamma = −0.13; negative gammas indicate that T-ACASI shifted responses toward *agreement* that preschool children suffer if their mothers work).

SUBPOPULATION DIFFERENCES IN THE IMPACT OF T-ACASI: HIERARCHICAL
LOG-LINEAR MODELS

Hierarchical log-linear models were used to examine subpopulation differences in the impact of T-ACASI on the one subjective measurement (endorsement of spanking) that showed a significant mode effect when the measurement was treated as a nominal—but not an ordinal—variable. Log-linear models were fit to the three-way distributions of response to the spanking question *by* survey mode *by* each sociodemographic variable taken separately. To test for an interaction (i.e., nonequivalence of the survey mode effect across categories of the sociodemographic variable), we fit a model that was constrained to fit all three two-way marginals. Failure to obtain an adequate fit for this model indicates that the observed distribution requires one or more three-way interaction terms to allow for nonequivalence(s) of the survey mode effect across subpopulations (i.e., across categories of the sociodemographic variable).

Evidence of statistically significant heterogeneity in the survey mode effect on responses to the spanking question was found for one sociodemographic variable—race (Black versus non-Black; likelihood ratio chi-square = 9.89, df = 3, $p = .020$). The three-way table for this tabulation is shown in table 3. When a saturated log-linear model was fit to this tabulation, it was found that variation between Blacks and non-Blacks in the impact of T-ACASI on the use of the "strongly disagree" category was largely responsible for this result. T-ACASI had almost no effect on the likelihood that *non-Black* respondents would say they "strongly disagree" that children sometimes need a "good, hard spanking" (21.7 percent versus 22.5 percent, OR = 0.95), while for *Black* respondents, T-ACASI decreased the odds of a strongly disagree response by a factor of 0.38 (5.6 percent in the T-ACASI condition versus 13.5 percent in T-IAQ).

Discussion

This research was intended to assess the overall impact of T-ACASI on measurements of subjective phenomena and variations in that impact, if any, across population subgroups. Our results indicate that T-ACASI does have a statistically reliable impact on survey measurements of a range of subjective phenomena.

Table 2. Results of Hypothesis Tests for Homogeneity of Estimated T-ACASI Mode Effects across Selected Sociodemographic Groups in Ordered Logit Models (Results Reported If *p*-value Was .10 or Less)

	T-ACASI interaction with							
	Education		Age		Baltimore versus national strata		Gender	
Subjective measurement	*p*	Variation in the T-ACASI effect	*p*	Variation in the T-ACASI effect	*p*	Variation in the T-ACASI effect	*p*	Variation in the T-ACASI effect
There is no real risk in occasional pot use	.03	Weakens with increased education[b]						
Preschool children suffer if mother works			.01	Weakens with age[d]	.03	Stronger in Baltimore[g]		
Much better if man is achiever outside home and woman takes care of home and family					.04	Stronger in Baltimore[h]		
Opinion about sex between adults of the same gender							.07	Stronger for males[j]
Compared to average person, how good looking are you?	.09	Weakens with increased education[a]	.04	Weakens with age[e]				

T-ACASI interaction with

Subjective measurement	Education		Age		Baltimore versus national strata		Gender	
	Variation in the T-ACASI effect	p	Variation in the T-ACASI effect	p	Variation in the T-ACASI effect	p	Variation in the T-ACASI effect	p
Preferred racial composition of neighborhood, if you could live anywhere (Black and White respondents only[a])	Weakens with increased education[c]	.03	Weakens with age[f]	.03	Stronger in Baltimore[i]	.07	Stronger for females[k]	.04

NOTE.—The table notes below provide measures of ordinal correlation between the binary variable identifying whether the respondent was interviewed by a human interviewer or T-ACASI and the ordinal correlation of the subjective measurement. Single binary variables represented the dichotomies: Black versus non-Black; female versus male; currently married or cohabiting versus other marital statuses; Baltimore versus U.S. sample strata; and ever versus never smoked marijuana. Multiple binary variables were used to represent the four categories of the population density and the six regions of the USA (see the footnotes to table 1 for list of categories). For years of education, we assigned values of 7 years for the categories "8th grade or less"; 11 years for "some high school"; 12 years for "finished high school or equivalent"; 13 years for "vocation, trade, or business school"; 14 years for "some college or two-year degree"; 16 years for "finished college with a four- or five-year degree"; and 17.5 for "postgraduate degree." The p-values reported above are derived from individual tests for each variable that compared ordered logit models that included variables representing the mode of interview (T-ACASI versus T-IAQ) and each sociodemographic factor (e.g., region of USA) taken separately for a model including these same variable plus variable(s) representing survey mode-by-sociodemographic interactions (e.g., region-by-mode).
[a] Gamma was −0.36 for respondents who did not graduate high school; −0.18 for those with some college; and −0.15 for college graduates.
[b] Gamma was −0.23 for respondents who did not graduate high school; −0.15 for high school graduates; −0.04 for those with some college; and 0.01 for college graduates.
[c] Gamma was 0.39 for respondents who did not graduate high school; 0.26 for high school graduates; 0.24 for those with some college; and 0.16 for college graduates.
[d] Gamma was −0.21 for respondents aged 18–25; −0.07 for those aged 26–35; and +0.07 for persons aged 36–45.
[e] Gamma was −0.31 for respondents aged 18–25; −0.24 for those aged 26–35; and −0.16 for persons aged 36–45.
[f] Gamma was 0.28 for respondents aged 18–25; 0.29 for those aged 26–35; and 0.12 for those aged 36–45.
[g] Gamma was 0.00 for the national sample stratum and −0.18 for the Baltimore stratum.
[h] Gamma was −0.02 for the national sample stratum and −0.18 for the Baltimore stratum.
[i] Gamma was 0.21 for the national sample stratum and 0.34 for the Baltimore stratum.
[j] Gamma for males was −0.17; it was −0.04 for females.
[k] Gamma for males was 0.12; it was 0.31 for females.

Table 3. Endorsement of Statement That Children Sometimes Need a "Good, Hard Spanking" by Mode of Interview and Race: Analysis of Unweighted Data from the NSBME

| Measurement and survey mode | Response to the spanking question | | | | | p^a | |
	Strongly agree	Agree	Disagree	Strongly disagree	N	Nominal	Ordinal
Black respondents							
T-ACASI	17.6%	38.0%	38.9%	5.6%	216	<.001	.189
T-IAQ	12.1%	46.8%	27.6%	13.5%	297		
Non-Black respondents							
T-ACASI	10.1%	32.9%	35.2%	21.7%	681	.087	.447
T-IAQ	6.9%	36.7%	33.9%	22.5%	843		

[a]p-values calculated for likelihood ratio chi-square tests and Mantel–Haenszel tests for linear association.

This is, perhaps, not a surprising outcome given the growing body of evidence documenting substantial divergences in survey measurements of nominally objective phenomena (e.g., reports of sexual and drug-use behaviors) made using T-ACASI versus those made using human telephone interviewers (Corkrey and Parkinson 2002; Lau, Tsui, and Wang 2003; Turner et al. 1996, 2005; Villarroel et al. 2006a). We were surprised, however, by the consistency of this impact. For all seven subjective measurements, we found statistically reliable divergences in the response distributions obtained in the T-ACASI and T-IAQ survey modes.

Furthermore, in six of the seven instances, we found a statistically significant or borderline shifting of response distributions along the ordinally scaled categories of our subjective measurements rather than a nondirectional movement of responses among response categories. T-ACASI responses were shifted (1) toward agreeing that men should be the "achiever" in the family; (2) toward agreeing that preschool children suffer if their mothers work; (3) toward saying that same-gender sex is wrong; (4) toward expressing a preference for living in racially homogeneous neighborhoods; and (5) toward agreeing that occasional marijuana use poses no risk. Similarly, while there was no overall directional impact of T-ACASI on the endorsement of corporal punishment for children, T-ACASI did increase the frequency of respondents "strongly agree[ing]" that children sometimes need a "good, hard spanking" (11.9 percent versus 8.2 percent, OR = 1.51, p = .005).

With the available data, one cannot be fully confident that the privacy afforded by T-ACASI allows respondents to express their "true feelings," and thus that the foregoing results suggest that "true" public opinion may be more supportive of traditional gender roles and corporal punishment and less supportive of integrated neighborhoods and same-gender sex than we would be led to believe by traditional interviewer-administered surveys. As with all measurements of

subjective phenomena, there is no simple criterion that can be employed for validation such as is done to assess bias in reporting of age, medical visits, voter registration, personal bankruptcies, and drunk driving convictions (Birdsong et al. 1992; Locander et al. 1976; Sawyer et al. 1989; Turner and Martin 1984, vol. 1, Chapters 1 and 5). Studies of the impact of audio-CASI and T-ACASI on reporting of sexual and drug use behaviors are, however, slowly producing evidence that suggests that the increased reporting of stigmatized and illicit behaviors obtained by these new technologies does, in fact, represent more accurate reporting.[9]

ALTERNATIVE EXPLANATIONS

Some readers—indeed many members of our research team—were surprised by our results. Our initial suspicion that T-ACASI might affect responses to these subjective measurements was consistently borne out. However, the pattern of our results raised a range of questions,[10] including: (1) Could other factors—such as differences in response rates in the T-ACASI versus T-IAQ conditions or order of presentation of response categories—explain these results? (2) Is it reasonable to assume that responses in the T-ACASI survey mode are more accurate indicators of the "true" subjective state of the respondent than T-IAQ measurements? (3) Why do T-ACASI respondents give more *conservative* answers to the three questions on sex roles and child rearing while give an apparently more *liberal* answer to the question of the risk of marijuana use? (4) Would the research team have predicted this pattern of results prior to viewing the results?

None of these questions can be definitively answered without further research, but we offer the following observations. First, the two most promising

9. So, for example, both the 2000 National Survey of Drug Use and Health (NSDUH) and the interviewer-administered condition in the NSBME national sample obtained almost no reports of (illicit) injection drug use in the prior year (population weighted estimates were unbelievably low: 0.1 percent for adults aged 18–49 in NSDUH and less than 0.05 percent for adults aged 18–45 interviewed in the T-IAQ condition of the NSBME; Turner et al. 2005, table 6). In contrast, the T-ACASI condition of the NSBME produced a population-weighted estimate of 1.5 percent. Similarly, the 1987 National Survey of Adolescent Males estimated that only 2.2 percent of young males had ever engaged in active or passive genital touching, oral sex, or anal sex with another male. (NSAM recruited a probability sample of males aged 15–19 in 1987 and made these measurements using paper SAQs). Reports made by adult men detailing their same-gender experiences during adolescence suggest that the prevalence of these behaviors should be two to four times higher than estimated by the 1987 NSAM—in the range of 4–9 percent (Turner et al. 1998b, note 13). The 1995 NSAM recruited a new probability sample of males aged 15–19 and administered the same questions in a randomized experiment testing the use of audio-CASI versus paper SAQs to make these measurements. In 1995, NSAM's paper SAQs yielded roughly the same population estimate as the 1987 NSAM (1.5 percent) while audio-CASI yielded an estimate (5.5 percent) in the expected range.
10. We thank the anonymous reviewers of a prior draft of this article for adding to the following list of questions.

Table 4. Preferred Racial Composition of Neighborhood by Interview Mode for Black and White Respondents: Analysis of Unweighted Data from the NSBME

Race and	Response				*p*	
interview mode	All or mostly Black	50-50	All or mostly White	*N*	Nominal	Ordinal
Black respondents						
T-ACASI	25.7%	67.8%	6.5%	214	.002	.002
T-IAQ	13.1%	78.5%	8.4%	275		
White respondents						
T-ACASI	0.7%	24.3%	75.1%	569	<.0001	<.0001
T-IAQ	0.9%	36.4%	62.7%	656		

NOTE.—Analysis collapses categories (1) all Black and mostly Black, and (2) all White and mostly White. Test for interaction compared ordered logit models that modeled neighborhood preference as a function of (1) race and survey mode, and (2) race, survey mode, and interaction term race-by-survey mode. Tests of fit of models reject the null hypothesis of no race mode interaction with $p < .0001$.

"alternative" methodological explanations are not well supported by the NS-BME data. The response rates in the T-ACASI condition were 6 (Baltimore) to 9 (USA) percentage points lower than those in the T-IAQ condition. Thus variation in the composition of the T-ACASI and T-IAQ samples could affect our outcomes. We found, however, that the introduction of statistical controls for nine sociodemographic variables did not markedly affect our conclusion about the impact of T-ACASI. Similarly, while the observed T-ACASI effects involved movement of respondents into the response category that was offered first, there is one piece of persuasive evidence that T-ACASI respondents were responding to the substantive meaning of questions and not the order of response categories. As table 4 shows, when asked about their preferences for neighborhoods, White respondents in the T-ACASI condition were shifted by 12 percentage points toward the *last-offered* response categories (all or mostly White: 75.1 percent in T-ACASI versus 62.7 percent in T-IAQ) while Black respondents were shifted by 12 percentage points toward the *first-offered* response categories (all or mostly Black: 25.7 percent in T-ACASI versus 13.1 percent in T-IAQ).

The results in table 4 also provide some basis for inferring that impression management lies behind the different responses obtained by T-ACASI and T-IAQ interviewing. Both Black and White respondents express more support for integration when interviewed by human interviewers and less support when interviewed by a T-ACASI computer. This result provides a basis for suspecting that responses in the T-ACASI survey mode are more accurate indicators of the "true" subjective state of the respondent than T-IAQ measurements. (This assumes, of course, that there is a social desirability bias that encourages

respondents *to report* favoring racial integration. Since fewer than 15 percent of respondents in either survey condition reported wanting to live in a neighborhood in which all the residents were of the same race, this is a reasonable assumption.)

We do not have a satisfying explanation for why T-ACASI elicits both more frequent agreement that there is no risk to occasional marijuana use and also more frequent agreement with traditional views of sex roles and child rearing. A shift toward selection of the first-offered response in T-ACASI could account for these results, but that would not account for the preceding "residential integration" example.

As noted earlier, the NSBME research team did not make explicit predictions about the impact of T-ACASI on these seven subjective measurements. The NSBME principal investigator reports, for example, that he expected that the question asking respondents to rate their physical attractiveness would show a shift toward "worse looking than average" in the more private T-ACASI survey mode. The actual NSBME result is the opposite of this expectation: respondents report themselves to be more attractive when interviewed by T-ACASI. With the clarity of hindsight, one might now claim that this result reflected a social norm against "boasting" about one's physical attractiveness. Such a facile reinterpretation should, however, invite skepticism.

Given the puzzles in the foregoing results, there is a clear need for replication of this research with additional experimental controls for factors such as order of response presentation that could be affecting our results. Expansion of the range of subjective measurements examined will also be helpful. We would also suggest that the research community should consider an approach to interpretation of mode effects suggested by Schuman (1982): what appear to be survey measurement artifacts can be evidence of important but hidden social facts. If technical factors (e.g., variation in the sample composition) can be ruled out, then the T-ACASI mode effects could be providing an indicator of social issues that are sensitive, the particular viewpoints that respondents are likely to hide from strangers, and the segments of the population that are most likely to do so. Findings from our examination of subpopulation differences in the T-ACASI mode effect in the NSBME are instructive in the latter instance.

SUBPOPULATION DIFFERENCES IN THE T-ACASI IMPACT

The impact of T-ACASI is not equivalent across all sections of the population. Less educated and younger respondents show the strongest mode effects, and the mode effect was also stronger in the sample stratum drawn from the city of Baltimore than in the national sample—the population of Baltimore City is predominantly minority, less educated, and poorer than the national population (U.S. Bureau of the Census 2007).

These results might be seen as supporting a general conclusion that the impact of human interviewers (versus T-ACASI computers) is most substantial

Table 5. Endorsement of Statement That There Is No Real Risk in Occasional Use of Marijuana by Mode of Interview and Race: Analysis of Unweighted Data from the NSBME

Measurement and survey mode	Response to marijuana risk question					p[a]	
	Strongly agree	Agree	Disagree	Strongly disagree	N	Nominal	Ordinal
Non-Black respondents							
T-ACASI	10.5%	21.9%	36.6%	31.0%	667	.061	.494
T-IAQ	7.0%	25.4%	36.8%	30.8%	834		
Black respondents							
T-ACASI	10.6%	15.0%	44.9%	29.5%	207	.002	.011
T-IAQ	7.1%	15.9%	31.9%	45.1%	295		

[a]*p*-values calculated for likelihood ratio chi-square tests and Mantel–Haenszel tests for linear association.

for more socially vulnerable or less self-confident segments of the population. This would include the young, the less educated, and the poor.[11] In considering these results, it is important to bear in mind the social context of these survey interviews. All respondents were informed that the study was being conducted by a major university and that the research was funded by the National Institutes of Health. Respondents in the interviewer-administered condition were then questioned by well-spoken, college-educated interviewers.[12] It is not unreasonable to suspect that the respondents who would be most likely to conceal unpopular attitudes in this context would be the youngest and least educated respondents.

Two other results could be seen as lending additional support to the proposed generalization. These results involve Black respondents and the differential impact of T-ACASI on the two survey measurements that might have legal consequences. First, as shown in table 5, the T-ACASI mode effect on marijuana measurements arose entirely from Black respondents. Treating the four response categories as an ordinal scale, we find that non-Black respondents gave statistically equivalent responses ($p = .494$) to human interviewers and to the T-ACASI computer. In contrast, for Black respondents, T-ACASI produced a statistically significant ($p = .011$) shift of responses toward agreement

11. It should be borne in mind, however, that these relationships held for only three of seven subjective measurements for both age and education. However, the direction of the mode effects was consistent in each instance; the impact of T-ACASI was stronger for younger and less educated respondents.

12. The survey unit has a requirement that all interviewers have at least some college education. Supervisors of the telephone survey facility estimate that about 60 percent of telephone interviewers were college graduates and the rest had some college education—split evenly among those currently pursuing degrees and those not currently in school (Anthony Roman, personal communication, March 3, 2008).

with the statement that there is little risk in occasional marijuana use. Indeed, 45 percent of Black respondents told human interviewers that they "strongly disagreed" that occasional marijuana use posed no risk while only 30 percent gave this same response when responding to the T-ACASI computer ($p < .001$). In contrast, 31 percent of non-Black respondents in *both* the T-ACASI and T-IAQ conditions said that they "strongly disagreed" that occasional marijuana use posed no risk.

A parallel divergence in T-ACASI's impact for Black respondents is seen for reporting strong disagreement with the statement that "It is sometimes necessary to discipline a child with a good, hard spanking" (see table 3). Human interviewers and T-ACASI were equally likely to elicit a "strongly disagree" response from non-Black respondents (22.5 percent versus 21.7 percent). However, Black respondents were 2.6 times more likely to say that they "strongly disagree" with this statement when interviewed by a human interviewer rather than a T-ACASI computer (13.5 percent versus 5.6 percent).[13]

If these substantial T-ACASI mode effects among Black respondents arise from greater sensitivity or perceived vulnerability among Black respondents to disclosure of favorable attitudes toward marijuana and corporal punishment, then one must ask if there is other evidence to suggest that is so. Unfortunately, there is none from the survey itself. However, heightened sensitivity or perceived vulnerability on the marijuana question would be consistent with claims that Blacks are arrested for marijuana possession at two and a half times the rate of Whites (Gettman/NORML 2000). Similarly, while "a good, hard spanking" need not be child abuse, it does occasionally lead to trouble with law enforcement or child welfare authorities, and some advocates argue that: "African-American children and families are significantly more likely to be torn apart by the local child welfare agency than any other racial group in the country—despite the fact that multiple national research studies prove that African-American families are not more likely to abuse or neglect their children" (Center for the Study of Social Policy 2004; see also Hill 2006).

While we believe that this interpretation is plausible, it is only one of a large number of interpretations that could be offered. We note, for example, that of the 24 interviewers who did the bulk of the NSBME interviews, 19 were White and only 4 were Black (Roman, personal communication, July 13, 2007). Black respondents were thus more likely to be interviewed by an interviewer of a different race than White respondents. While we do not know if the White NSBME interviewers were perceived as "White" by respondents, one might conjecture that the foregoing pattern of results could have arisen from a greater willingness of respondents to disclose their "true" opinions on these

---

13. This heterogeneity by race in the impact of survey mode on "strongly disagreeing" with spanking is statistically reliable. Dichotomizing the response variable into "strongly disagree" (1) or not (0), the test for a three-way interaction yields the following: likelihood ratio chi-square = 7.11, df = 1, $p = .008$.

topics in interviews conducted by same-race human interviewers and by T-ACASI and a lesser willingness to disclose these same opinions to different-race interviewers.

Conclusion

These results suggest five major conclusions. First, the response distributions for all subjective measurements obtained by T-ACASI diverge from those obtained by human telephone interviewers. Second, for six of our seven ordinal-scaled measurements, this divergence involved shifting responses along the ordinal scale and not a nondirectional redistribution among response categories. Third, when interviewed by T-ACASI, respondents were more supportive of traditional gender roles and corporal punishment, less supportive of integrated neighborhoods and same-gender sex, and more likely to agree that occasional marijuana use is harmless and to describe themselves as attractive. Fourth, although the evidence is not entirely consistent, it appears that the impact of T-ACASI may increase with the social vulnerability of the populations surveyed. Finally, replication and extension of this research should be a priority for those concerned about subjective measurements.

Appendix

QUESTION WORDINGS

1–4. Now we will present four statements. After each one, please tell us whether you strongly agree, agree, disagree, or strongly disagree with it.

It is sometimes necessary to discipline a child with a good, hard spanking.
There is no real risk associated with the occasional use of marijuana.
A preschool child is likely to suffer if his or her mother works.
It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family.

5. We would like to know your opinion about sexual relations between two adults of the same sex—do you think it is: always wrong, almost always wrong, wrong only sometimes, or not wrong at all.

6. Compared to the average person, how good looking do you think you are? Much better looking, somewhat better looking, average, somewhat worse looking, much worse looking. *(Note. Only four respondents described themselves as "much worse looking." In our analyses, this response category was combined with "somewhat worse looking.")*

7. If you could find the housing that you would want and like, would you rather live in a neighborhood that is all Black; mostly Black; half Black, half White; mostly White; or all White? *(Our analyses of this measurement*

is restricted to respondents who said their race was White or Black. Their responses were recoded to: all the same race, mostly the same race, half and half, mostly the other race, or all the other race. Since only nine respondents reported preferring neighborhoods that were all of a different race, our analyses collapsed the "mostly different" and "all different" response categories.)

References

American Association for Public Opinion Research (AAPOR). 2006. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates*. 4th ed. Lenexa, KS: AAPOR.

Anderson, John E., William D. Mosher, and Anjani Chandra. 2006. "Measuring HIV risk in the U.S. Population Aged 15–44: Results from Cycle 6 of the National Survey of Family Growth."*Advanced Data from Vital and Health Statistics* 23:1–27.

Birdsong, William H., Andrea A. Lash, Sarah Thayer, Eugene Kumekawa, and Charles E. Becker. 1992. "The Validity of Study Group Assignments Based on Occupational Histories Obtained from Questionnaires." *Journal of Occupational Medicine* 34:940–45.

Brener, Nancy D., Danice K. Eaton, Laura Kann, Jo Anne Grunbaum, Lori A. Gross, Tonja M. Kyle, and James G. Ross. 2006. "The Association of Survey Setting and Mode with Self-reported Health Risk Behaviors among High School Students." *Public Opinion Quarterly* 70: 354–74.

Center for the Study of Social Policy. 2004. "The Race and Child Welfare Project: Social Fact Sheet 1, Basic Facts on Disproportionate Representation of African Americans in the Foster Care System." Available at http://www.cssp.org/uploadFiles/factSheet1.pdf.

Cooley, Phillip C., Heather G. Miller, James N. Gribble, and Charles F. Turner. 2000. "Automating Telephone Surveys: Using T-ACASI to Obtain Data on Sensitive Topics." *Computers and Human Behavior* 16:1–11.

Corkrey, Ross, and Lynne Parkinson. 2002. "A Comparison of Four Computer-based Telephone Interviewing Methods: Getting Answers to Sensitive Questions." *Behavior Research Methods: Instruments and Computers* 34:354–63.

Currivan, Douglas, Amy L. Nyman, Charles F. Turner, and Lois Biener. 2004. "Does Telephone Audio Computer-Assisted Self-interviewing Improve the Accuracy of Prevalence Estimates of Youth Smoking?" *Public Opinion Quarterly* 68:542–64.

Davis, James A., Tom W. Smith, and Peter V. Marsden. 2002. *General Social Surveys, 1972–2002: Cumulative Codebook*. Chicago, IL: National Opinion Research Center.

DesJarlais, Don C., Denise Paone, Judith Milliken, Charles F. Turner, Heather G. Miller, James N. Gribble, Quihu Shi, Holly Hagan, Samuel R. Friedman. 1999. "Using Audio-Computer Inter-viewing to Measure Risk Behavior for HIV among Injecting Drug Users: A Quasi-randomised Trial." *Lancet* 353:1657–61.

Duffy John C., and Jenniffer J. Waterton. 1984. "Under-reporting of Alcohol Consumption in Sam-ple Surveys: The Effect of Computer Interviewing in Fieldwork." *British Journal of Addiction* 79:303–308.

Epstein, Joan F. 2002. "Substance Dependence, Abuse, and Treatment: Findings from the 2000 National Household Survey on Drug Abuse." NHSDA Series A-16, DHHS Publication No. SMA 02-3642. Substance Abuse and Mental Health Services Administration, Office of Applied Studies, Rockville, MD.

Fay, Robert E., Charles F. Turner, Albert D. Klassen, and John H. Gagnon. 1989. "Preva-lence and Patterns of Same-Gender Sexual Contact among Men." *Science* 243:338–48.

Gensys Sampling Systems. 2002. "Gensys Methodologies." Available at http://www.genesys-sampling.com/reference/genmeth.htm.

Gettman, Jon, and the NORML Foundation. 2000. "United States Marijuana Arrests: Part Two. Racial Differences in Drug Arrests." Available at http://www.norml.org/index.cfm?Group_ID= 5328#abstract (accessed July 2007).

Goodman, Leo A. 1968. "The Analysis of Cross-classified Data: Independence, Quasi-independence, and Interactions in Contingency Tables." *Journal of the American Statistical Association* 63:1091–131.

Gribble, James N., Heather G. Miller, Joseph A. Catania, Lance Pollack, and Charles F. Turner. 2000. "The Impact of T-ACASI Interviewing on Reported Drug Use among Men Who Have Sex with Men." *Substance Use and Misuse* 35:869–90.

Hill, Robert B. 2006. *Synthesis of Research on Disproportionality in Child Welfare: An Update*. Baltimore, MD: Annie E. Casey Foundation. http://www.cssp.org/uploadFiles/ Disproportionality_Paper_Bob_Hill.pdf.

Kobak, Kenneth A, Leslie H. Taylor, Susan L. Dottl, John H. Greist, James W. Jefferson, Diane Burroughs, Julie M. Mantle, David J. Katzelnick, Randal Norton, Henry J. Henk, and Ronald C. Serlin. 1997. "A Computer-administered Telephone Interview to Identify Mental Disorders." *Journal of the American Medical Association* 278:905–10.

Lau, Joseph T.F., Hi-Yi Tsui, and Qing-Song Wang. 2003. "Effects of Two Telephone Survey Methods on the Level of Reported Risk Behaviors." *Sexually Transmitted Infections* 79: 325–31.

Locander, William, Seymour Sudman, and Norman Bradburn. 1976. "An Investigation of Interview Method, Threat and Response Distortion." *Journal of the American Statistical Association* 71:269–75.

Long, J. Scott, and Jeremy Freese. 2001. *Regression Models for Categorical Dependent Variables Using Stata*. College Station, TX: Stata Press.

Millard, Richard W., and Joseph R. Carver. 1999. "Cross-sectional Comparison of Live and Interactive Voice Recognition Administration of the SF-12 Health Status Survey." *American Journal of Managed Care* 5:153–59.

Miller, Heather G., Charles F. Turner, and Lincoln Moses, eds. 1990. *AIDS: The Second Decade*. Washington, DC: National Academy Press.

Moskowitz, Joel M. 2004. "Assessment of Cigarette Smoking and Smoking Susceptibility among Youth: Telephone Computer-assisted Self-interviews versus Computer-assisted Telephone Interviews." *Public Opinion Quarterly* 68:565–87.

O'Reilly, James, Michael Hubbard, Judith Lessler, Paul Biemer, and Charles F. Turner. 1994. "Audio and Video Computer-assisted Self-interviewing: Preliminary Tests of New Technologies for Data Collection." *Journal of Official Statistics* 10:197–214.

Roman, Anthony M. 2000. "Survey of Aids and Other Social Issues: Field Report." Unpublished manuscript, Center for Survey Research, University of Massachusetts, Boston.

Sawyer, John A., Jo Anne Earp, Robert H. Fletcher, Fedora F. Daye, and Tonja M. Wynn. 1989. "Accuracy of Women's Self-Report of Their Last Pap Smear." *American Journal of Public Health* 79:1036–37.

Schuman, Howard. 1982. "Artifacts Are in the Mind of the Beholder." *American Sociologist* 17:21–28.

Searles JS, Perrine MW, Mundt JC, Helzer JE. 1995. "Self-report of Drinking Using Touch-Tone Telephone: Extending the Limits of Reliable Daily Contact." *Journal of Studies on Alcohol* 56:375–82.

Sonenstein, Freya L., Leighton Ku, Laura D. Lindberg, Charles F. Turner, and Joseph H. Pleck. 1998. "Changes in Sexual Behavior and Condom Use among Teenage Men: 1988 to 1995." *American Journal of Public Health* 88:956–59.

SPSS, Inc. 1993. *SPSS Statistical Software*. Version 6.0.1. Chicago, IL: SPSS Inc.

Stata Corporation. 2003. *Stata Version 8/SE for Windows*. College Station, TX: Stata Corporation.

Sudman, Seymour, Andrew M. Greeley, and Leonard J. Pinto. 1965. "The Effectiveness of Self-administered Questionnaires." *Journal of Marketing Research* 2:293–97.

Tourangeau, Roger, Darby Miller Steiger, and David Wilson. 2002. "Self-administered Questions by Telephone: Evaluating Interactive Voice Response." *Public Opinion Quarterly* 66:265–78.

Tourangeau, Roger, and Tom W. Smith. 1996. "Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context." *Public Opinion Quarterly* 60:275–304.

Turner, Charles F., and Amy R. Sheon. 1994. "Behavioral Studies Relevant to Vaccine Trial Preparation: An Introduction." *AIDS Research and Human Retroviruses* 10:S273–6.

Turner, Charles F., Barbara H. Forsyth, James O'Reilly, Phillip C. Cooley, Timothy K. Smith, Susan M. Rogers, and Heather G. Miller. 1998a. "Automated Self-interviewing and the Survey Measurement of Sensitive Behaviors." In *Computer-Assisted Survey Information Collection*, eds. Mick Couper, Reginald P. Baker, Jelke Bethlehem, Cynthia Z. F. Clark, Jean Martin, William L. Nicholls, and James O'Reilly, pp. 455–73. New York: Wiley.

Turner, Charles F., and Elizabeth Martin, eds. 1984. *Surveying Subjective Phenomena*, vols. 1 and 2. New York: Russell Sage.

Turner, Charles F., Heather G. Miller, James A. Catania, Phillip C. Cooley, Paul Biemer, and James R. Chromy. 1995. "Improving AIDS Behavioral Studies Using T-ACASI." Unpublished research proposal to the National Institutes of Health for grant R01-MH056318. Research Triangle Institute, Washington, DC.

Turner, Charles F., Heather G. Miller, Timothy K. Smith, Phillip C. Cooley, and Susan M. Rogers. 1996a. "Telephone Audio Computer-Assisted Self-interviewing (T-ACASI) and Survey Measurements of Sensitive Behaviors: Preliminary Results." In *Survey and Statistical Computing 1996*, eds. Randy Banks, Joan Fairgrieve, Laurange Gerrard, Terry Orchard, Clive Payne, and Andrew Westlake, pp. 121–30. Chesham, Buckinghamshire: Association for Survey Computing.

Turner, Charles F., Judith Lessler, and James Devore. 1992. "Effects of Mode of Administration and Wording on Reporting of Drug Use." In *Survey Measurement of Drug Use: Methodological Issues*, eds. Charles F. Turner, Judith Lessler, and Joseph C. Gfroerer, pp. 177–219. DHHS Publication no. 92-1929. Washington, DC: Government Printing Office.

Turner, Charles F., Judith Lessler, and Joseph Gfroerer. 1992. "Future Directions for Research and Practice." In *Survey Measurement of Drug Use: Methodological Issues*, eds. Charles F. Turner, Judith Lessler, and Joseph C. Gfroerer, pp. 299–306. DHHS Publication no. 92-1929. Washington, DC: Government Printing Office.

Turner, Charles F., Leighton Ku, Freya L. Sonenstein, and Joseph H. Pleck. 1996b. Impact of Audio-CASI on Bias in Reporting of Male-Male Sexual Contacts. In *Health Survey Research Methods*, ed. Richard Warnecke. Hyattsville, MD: National Center for Health Statistics.

Turner, Charles F., Leighton Ku, Susan M. Rogers, Laura D. Lindberg, Joseph H. Pleck, and Freya L. Sonenstein. 1998b. "Adolescent Sexual Behavior, Drug Use, and Violence: Increased Reporting with Computer Survey Technology." *Science* 280:867–73.

Turner, Charles F., Maria A. Villarroel, Susan M. Rogers, Elizabeth E. Eggleston, Laxminarayana Ganapathi, Anthony M. Roman, and Alia A. Al-Tayyib. 2005. "Reducing Bias in Telephone Survey Estimates of the Prevalence of Drug Use: A Randomized Trial of Telephone Audio-CASI." *Addiction* 100:1432–44.

Turner, Charles F., Susan M. Rogers, Heather G. Miller, William C. Miller, James N. Gribble, James R. Chromy, Peter A. Leone, Phillip C. Cooley, Thomas C. Quinn, and Jonathan Zenilman. 2002. "Untreated Gonococcal and Chlamydial Infection in a Probability Sample of Adults." *Journal of the American Medical Association* 287:726–33.

U.S. Bureau of the Census. 2007. *State and County Quick Facts (Baltimore City, Maryland)*. Washington, DC: U.S. Bureau of the Census. Available at http://quickfacts.census.gov/qfd/states/24/2404000.html.

Villarroel, Maria A., Charles F. Turner, Elizabeth E. Eggleston, Alia A. Al-Tayyib, Susan M. Rogers, Anthony M. Roman, Phillip C. Cooley, and Harper Gordek. 2006a. "Same-Gender Sex in the USA: Impact of T-ACASI on Prevalence Estimates." *Public Opinion Quarterly* 70:166–96.

Villarroel, Maria. A., Charles F. Turner, Elizabeth E. Eggleston, Alia A. Al-Tayyib, Susan M. Rogers, Anthony M. Roman, Phillip C. Cooley, and Harpur Gordek. 2006b. "Sample Design and Demographic Characteristics of Respondents in the 1999-2000 National STD and Behavior Measurement Experiment (NSBME)." *Technical Papers in Health and Behavior Measurement, No. 69*, Washington, DC: Program in Health and Behavior Measurement, Research Triangle Institute, 2006. Available at http://soc.qc.edu/turner/Reprints/TechPDFs/69_NSBMEsample.pdf.

Villarroel, Maria A., Charles F. Turner, Susan M. Rogers, Anthony M. Roman, Phillip C. Cooley, Allyna B. Steinberg, Elizabeth E. Eggleston, and James R. Chromy. 2008. "T-ACASI Reduces Bias in STD Measurements: The National STD and Behavior Measurement Experiment." *Sexually Transmitted Diseases* 35:499–506.