Sensitive Questions in Surveys

Roger Tourangeau
University of Maryland and University of Michigan

Ting Yan
University of Michigan

Psychologists have worried about the distortions introduced into standardized personality measures by social desirability bias. Survey researchers have had similar concerns about the accuracy of survey reports about such topics as illicit drug use, abortion, and sexual behavior. The article reviews the research done by survey methodologists on reporting errors in surveys on sensitive topics, noting parallels and differences from the psychological literature on social desirability. The findings from the survey studies suggest that misreporting about sensitive topics is quite common and that it is largely situational. The extent of misreporting depends on whether the respondent has anything embarrassing to report and on design features of the survey. The survey evidence also indicates that misreporting on sensitive topics is a more or less motivated process in which respondents edit the information they report to avoid embarrassing themselves in the presence of an interviewer or to avoid repercussions from third parties.

Keywords: sensitive questions, mode effects, measurement error, social desirability

Over the last 30 years or so, national surveys have delved into increasingly sensitive topics. To cite one example, since 1971, the federal government has sponsored a series of recurring studies to estimate the prevalence of illicit drug use, originally the National Survey of Drug Abuse, later the National Household Survey of Drug Abuse, and currently the National Survey on Drug Use and Health. Other surveys ask national samples of women whether they have ever had an abortion or ask samples of adults whether they voted in the most recent election. An important question about such surveys is whether respondents answer the questions truthfully. Methodological research on the accuracy of reports in surveys about illicit drug use and other sensitive topics, which we review in this article, suggests that misreporting is a major source of error, more specifically of bias, in the estimates derived from these surveys. To cite just one line of research, Tourangeau and Yan (in press) reviewed studies that compared self-reports about illicit drug use with results from urinalyses and found that some 30%–70% of those who test positive for cocaine or opiates deny having used drugs recently. The urinalyses have very low false positive rates (see, e.g., Wish, Hoffman, & Nemes, 1997), so those deniers who test positive are virtually all misreporting.

Most of the studies on the accuracy of drug reports involve surveys of special populations (such as enrollees in drug treatment programs or arrestees in jail), but similar results have been found for other sensitive topics in samples of the general population. For instance, one study compared survey reports about abortion from respondents to the National Survey of Family Growth (NSFG) with data from abortion clinics (Fu, Darroch, Henshaw, & Kolb, 1998). The NSFG reports were from a national sample of women between the ages of 15 and 44. Both the survey reports from the NSFG and the provider reports permit estimates of the total number of abortions performed in the U.S. during a given year. The results indicated that only about 52% of the abortions are reported in the survey. (Unlike the studies on drug reporting, the study by Fu et al., 1998, compared the survey reports against more accurate data in the aggregate rather than at the individual level.) Another study (Belli, Traugott, & Beckmann, 2001) compared individual survey reports about voting from the American National Election Studies with voting records; the study found that more than 20% of the nonvoters reported in the survey that they had voted. National surveys are often designed to yield standard errors (reflecting the error due to sampling) that are 1% of the survey estimates or less. Clearly, the reporting errors on these topics produce biases that are many times larger than that and may well be the major source of error in the national estimates.

In this article, we summarize the main findings regarding sensitive questions in surveys. In the narrative portions of our review, we rely heavily on earlier attempts to summarize the vast literature on sensitive questions (including several meta-analyses). We also report three new meta-analyses on several topics for which the empirical evidence is somewhat mixed. Sensitive questions is a broad category that encompasses not only questions that trigger social desirability concerns but also those that are seen as intrusive by the respondents or that raise concerns about the possible repercussions of disclosing the information. Possessing cocaine is not just socially undesirable; it is illegal, and people may misreport in a drug survey to avoid legal consequences rather than merely to avoid creating an unfavorable impression. Thus, the first issue we deal with is the concept of sensitive questions and its relation to the

concept of social desirability. Next, we discuss how survey respondents seem to cope with such questions. Apart from misreporting, those who are selected for a survey on a sensitive topic can simply decline to take part in the survey (assuming they know what the topic is), or they can take part but refuse to answer the sensitive questions. We review the evidence on the relation between question sensitivity and both forms of nonresponse as well as the evidence on misreporting about such topics. Several components of the survey design seem to affect how respondents deal with sensitive questions; this is the subject of the third major section of the article. On the basis of this evidence, we discuss the question of whether misreporting in response to sensitive questions is deliberate and whether the process leading to such misreports is nonetheless partly automatic. Because there has been systematic research on the reporting of illicit drug use to support the design of the federal drug surveys, we draw heavily on studies on drug use reporting in this review, supplementing these findings with related findings on other sensitive topics.

What Are Sensitive Questions?

Survey questions about drug use, sexual behaviors, voting, and income are usually considered sensitive; they tend to produce comparatively higher nonresponse rates or larger measurement error in responses than questions on other topics. What is it about these questions that make them sensitive? Unfortunately, the survey methods literature provides no clear answers. Tourangeau, Rips, and Rasinski (2000) argued that there are three distinct meanings of the concept of "sensitivity" in the survey literature.

Intrusiveness and the Threat of Disclosure

The first meaning of the term is that the questions themselves are seen as *intrusive*. Questions that are sensitive in this sense touch on "taboo" topics, topics that are inappropriate in everyday conversation or out of bounds for the government to ask. They are seen as an invasion of privacy, regardless of what the correct answer for the respondent is. This meaning of sensitivity is largely determined by the content of the question rather than by situational factors such as where the question is asked or to whom it is addressed. Questions asking about income or the respondent's religion may fall into this category; respondents may feel that such questions are simply none of the researcher's business. Questions in this category risk offending all respondents, regardless of their status on the variable in question.

The second meaning involves the *threat of disclosure*, that is, concerns about the possible consequences of giving a truthful answer should the information become known to a third party. A question is sensitive in this second sense if it raises fears about the likelihood or consequences of disclosure of the answers to agencies or individuals not directly involved in the survey. For example, a question about use of marijuana is sensitive to teenagers when their parents might overhear their answers, but it is not so sensitive when they answer the same question in a group setting with their peers. Respondents vary in how much they worry about the confidentiality of their responses, in part based on whether they have anything to hide. In addition, even though surveys routinely offer assurances of confidentiality guaranteeing nondisclosure, survey respondents do not always seem to believe these assur-

ances, so concerns about disclosure may still be an important factor in the misreporting of illegal or socially undesirable behaviors (Singer & Presser, in press; but see also Singer, von Thurn, & Miller, 1995).

Sensitivity and Social Desirability

The last meaning of question sensitivity, closely related to the traditional concept of social desirability, is the extent to which a question elicits answers that are socially unacceptable or *socially undesirable* (Tourangeau et al., 2000). This conception of sensitivity presupposes that there are clear social norms regarding a given behavior or attitude; answers reporting behaviors or attitudes that conform to the norms are deemed socially desirable, and those that report deviations from the norms are considered socially undesirable. For instance, one general norm is that citizens should carry out their civic obligations, such as voting in presidential elections. As a result, in most settings, admitting to being a nonvoter is a socially undesirable response. A question is sensitive when it asks for a socially undesirable answer, when it asks, in effect, that the respondent admit he or she has violated a social norm. Sensitivity in this sense is largely determined by the respondents' potential answers to the survey question; a question about voting is not sensitive for a respondent who voted.[1] Social desirability concerns can be seen as a special case of the threat of disclosure, involving a specific type of interpersonal consequence of revealing information in a survey—social disapproval.

The literature on social desirability is voluminous, and it features divergent conceptualizations and operationalizations of the notion of socially desirable responding (DeMaio, 1984). One fundamental difference among the different approaches lies in whether they treat socially desirable responding as a stable personality characteristic or a temporary social strategy (DeMaio, 1984). The view that socially desirable responding is, at least in part, a personality trait underlies psychologists' early attempts to develop various social desirability scales. Though some of these efforts (e.g., Edwards, 1957; Philips & Clancy, 1970, 1972) recognize the possibility that social desirability is a property of the items rather than (or as well as) of the respondents, many of them treat socially desirable responding as a stable personality characteristic (e.g., Crowne & Marlowe, 1964; Schuessler, Hittle, & Cardascia, 1978). By contrast, survey researchers have tended to view socially desirable responding as a response strategy reflecting the sensitivity of specific items for specific individuals; thus, Sudman and Bradburn (1974) had interviewers rate the social

---

[1] In addition, the relevant norms may vary across social classes or subcultures within a society. T. Johnson and van der Vijver (2002) provided a useful discussion of cultural differences in socially desirable responding. When there is such variation in norms, the bias induced by socially desirable responding may distort the observed associations between the behavior in question and the characteristics of the respondents, besides affecting estimates of overall means or proportions. For instance, the norm of voting is probably stronger among those with high levels of education than among those with less education. As a result, highly educated respondents are both more likely to vote and more likely to misreport if they did not vote than are respondents with less education. This differential misreporting by education will yield an overestimate of the strength of the relationship between education and voting.

desirability of potential answers to specific survey questions. Paulhus's (2002) work encompasses both viewpoints, making a distinction between socially desirable responding as a response *style* (a bias that is "consistent across time and questionnaires"; Paulhus, 2002, p. 49) and as a response *set* (a short-lived bias "attributable to some temporary distraction or motivation"; Paulhus, 2002, p. 49).

A general weakness with scales designed to measure socially desirable responding is that they lack "true" scores, making it difficult or impossible to distinguish among (a) respondents who are actually highly compliant with social norms, (b) those who have a sincere but inflated view of themselves, and (c) those who are deliberately trying to make a favorable impression by falsely reporting positive things about themselves. Bradburn, Sudman, and Associates (1979, see chap. 6) argued that the social desirability scores derived from the Marlowe–Crowne (MC) items (Crowne & Marlowe, 1964) largely reflect real differences in behaviors, or the first possibility we distinguish above:

> We consider MC scores to indicate personality traits . . . MC scores [vary] . . . not because respondents are manipulating the image they present in the interview situation, but because persons with high scores have different life experiences and behave differently from persons with lower scores. (p. 103)

As an empirical matter, factor analyses of measures of socially desirable responding generally reveal two underlying factors, dubbed the Alpha and Gamma factors by Wiggins (Wiggins, 1964; see Paulhus, 2002, for a review). Paulhus' early work (Paulhus, 1984) reflected these findings, dividing social desirability into two components: self-deception (corresponding to Wiggins' Alpha factor and to the second of the possibilities we distinguish above) and impression management (corresponding to Gamma and the third possibility above). (For related views, see Messick, 1991; Sackheim & Gur, 1978.) The Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1984) provides separate scores for the two components.

Paulhus's (2002) later work went even further, distinguishing four forms of socially desirable responding. Two of them involve what he calls "egoistic bias" (p. 63), or having an inflated opinion of one's social and intellectual status. This can take the form of self-deceptive enhancement (that is, sincerely, but erroneously, claiming positive characteristics for oneself) or more strategic agency management (bragging or self-promotion). The other two forms of socially desirable responding are based on "moralistic bias" (p. 63), an exaggerated sense of one's moral qualities. According to Paulhus, self-deceptive denial is a relatively unconscious tendency to deny one's faults; communion (or relationship) management is more strategic, involving deliberately minimizing one's mistakes by making excuses and executing other damage control maneuvers. More generally, Paulhus (2002) argued that we tailor our impression management tactics to our goals and to the situations we find ourselves in. Self-promotion is useful in landing a job; making excuses helps in avoiding conflicts with a spouse.

Assessing the Sensitivity of Survey Items

How have survey researchers attempted to measure socially desirability or sensitivity more generally? In an early attempt, Sudman and Bradburn (1974) asked coders to rate the social desirability of the answers to each of a set of survey questions on a 3-point scale (no possibility, some possibility, or a strong possibility of a socially desirable answer). The coders did not receive detailed instructions about how to do this, though they were told to be "conservative and to code 'strong possibility' only for questions that have figured prominently in concern over socially desirable answers" (Sudman & Bradburn, 1974, p. 43). Apart from its vagueness, a drawback of this approach is its inability to detect any variability in the respondents' assessments of the desirability of the question content (see DeMaio, 1984, for other criticisms).

In a later study, Bradburn et al. (1979) used a couple of different approaches to get direct respondent ratings of the sensitivity of different survey questions. They asked respondents to identify questions that they felt were "too personal." They also asked respondents whether each question would make "most people" very uneasy, somewhat uneasy, or not uneasy at all. They then combined responses to these questions to form an "acute anxiety" scale, used to measure the degree of threat posed by the questions. Questions about having a library card and voting in the past election fell at the low end of this scale, whereas questions about bankruptcy and traffic violations were at the high end. Bradburn et al. subsequently carried out a national survey that asked respondents to judge how uneasy various sections of the questionnaire would make most people and ranked the topics according to the percentage of respondents who reported that most people would be "very uneasy" or "moderately uneasy" about the topic. The order was quite consistent with the researchers' own judgments, with masturbation rated the most disturbing topic and sports activities the least (Bradburn et al., 1979, Table 5).

In developing social desirability scales, psychological researchers have often contrasted groups of participants given different instructions about how to respond. In these studies, the participants are randomly assigned to one of two groups. Members of one group are given standard instructions to answer according to whether each statement applies to them; those in the other group are instructed to answer in a socially desirable fashion regardless of whether the statement actually applies to them (that is, they are told to "fake good"; Wiggins, 1959). The items that discriminate most sharply between the two groups are considered to be the best candidates for measuring social desirability. A variation on this method asks one group of respondents to "fake bad" (that is, to answer in a socially undesirable manner) and the other to "fake good." Holbrook, Green, and Krosnick (2003) used this method to identify five items prone to socially desirable responding in the 1982 American National Election Study.

If the consequences of sensitivity were clear enough, it might be possible to measure question sensitivity indirectly. For example, if sensitive questions consistently led to high item nonresponse rates, this might be a way to identify sensitive questions. Income questions and questions about financial assets are usually considered to be sensitive, partly because they typically yield very high rates of missing data—rates as high as 20%–40% have been found across surveys (Juster & Smith, 1997; Moore, Stinson, & Welniak, 1999). Similarly, if sensitive questions trigger a relatively controlled process in which respondents edit their answers, it should take more time to answer sensitive questions than equally demanding nonsensitive ones. Holtgraves (2004) found that response times got longer when the introduction to the questions heightened social desirability concerns, a finding consistent with the operation of an

editing process. Still, other factors can produce high rates of missing data (see Beatty & Herrmann, 2002, for a discussion) or long response times (Bassili, 1996), so these are at best indirect indicators of question sensitivity.

Consequences of Asking Sensitive Questions

Sensitive questions are thought to affect three important survey outcomes: (a) overall, or unit, response rates (that is, the percentage of sample members who take part in the survey), (b) item nonresponse rates (the percentage of respondents who agree to participate in the survey but who decline to respond to a particular item), (c) and response accuracy (the percentage of respondents who answer the questions truthfully). Sensitive questions are suspected of causing problems on all three fronts, lowering overall and item response rates and reducing accuracy as well.

Unit Response Rates

Many survey researchers believe that sensitive topics are a serious hurdle for achieving high unit response rates (e.g., Catania, Gibson, Chitwood, & Coates, 1990), although the evidence for this viewpoint is not overwhelming. Survey texts often recommend that questionnaire designers keep sensitive questions to the end of a survey so as to minimize the risk of one specific form of nonresponse—break-offs, or respondents quitting the survey part way through the questionnaire (Sudman & Bradburn, 1982).

Despite these beliefs, most empirical research on the effects of the topic on unit response rates has focused on topic saliency or interest rather than topic sensitivity (Groves & Couper, 1998; Groves, Presser, & Dipko, 2004; Groves, Singer, & Corning, 2000). Meta-analytic results point to topic saliency as a major determinant of response rates (e.g., Heberlein & Baumgartner, 1978); only one study (Cook, Heath, & Thompson, 2000) has attempted to isolate the effect of topic sensitivity. In examining response rates to Web surveys, Cook et al. (2000) showed that topic sensitivity was negatively related to response rates, but, though the effect was in the expected direction, it was not statistically significant.

Respondents may be reluctant to report sensitive information in surveys partly because they are worried that the information may be accessible to third parties. Almost all the work on concerns about confidentiality has examined attitudes toward the U.S. census. For example, a series of studies carried out by Singer and her colleagues prior to Census 2000 suggests that many people have serious misconceptions about how the census data are used (Singer, van Hoewyk, & Neugebauer, 2003; Singer, Van Hoewyk, & Tourangeau, 2001). Close to half of those surveyed thought that other government agencies had access to names, addresses, and other information gathered in the census. (In fact, the confidentiality of the census data as well as the data collected in other federal surveys is strictly protected by law; data collected in other surveys may not be so well protected.) People with higher levels of concern about the confidentiality of the census data were less likely to return their census forms in the 1990 and 2000 censuses (Couper, Singer, & Kulka, 1998; Singer et al., 2003; Singer, Mathiowetz, & Couper, 1993). Further evidence that confidentiality concerns can affect willingness to respond at all comes from experiments conducted by the U.S. Census Bureau that asked respondents to

provide their Social Security Numbers in a mail survey; this lowered the unit response rate and raised the level of missing data among those who did mail back the questionnaire (Dillman, Sinclair, & Clark, 1993; see also Guarino, Hill, & Woltman, 2001). Similarly, an experiment by Junn (2001) showed that when confidentiality issues were made salient by questions about privacy, respondents were less likely to answer the detailed questions on the census long form, resulting in a higher level of missing data.

Thus, concerns about confidentiality do seem to contribute both to unit and item nonresponse. To address these concerns, most federal surveys include assurances about the confidentiality of the data. A meta-analysis by Singer et al. (1995) indicated that these assurances generally boost overall response rates and item response rates to the sensitive questions (see also Berman, McCombs, & Boruch, 1977). Still, when the data being requested are not all that sensitive, elaborate confidentiality assurances can backfire, lowering overall response rates (Singer, Hippler, & Schwarz, 1992).

Item Nonresponse

Even after respondents agree to participate in a survey, they still have the option to decline to answer specific items. Many survey researchers believe that the item nonresponse rate increases with question sensitivity, but we are unaware of any studies that systematically examine this hypothesis. Table 1 displays item nonresponse rates for a few questions taken from the NSFG Cycle 6 Female Questionnaire. The items were administered to a national sample of women who were from 15 to 44 years old. Most of the items in the table are from the interviewer-administered portion of the questionnaire; the rest are from the computer-administered portion. This is a survey that includes questions thought to vary widely in sensitivity, ranging from relatively innocuous sociodemographic items to detailed questions about sexual behavior. It seems apparent from the table that question sensitivity has some positive relation to item nonresponse rates: The lowest rate of missing data is for the least sensitive item (on the highest grade completed), and the highest rate is for total income question. But the absolute differences across items are not very dramatic, and lacking measures of the sensitivity of each item, it is difficult to assess the strength of the overall relationship between item nonresponse and question sensitivity.

Table 1

Item Nonresponse Rates for the National Survey of Family Growth Cycle 6 Female Questionnaire, by Item

Item	Mode of administration	%
Total household income	ACASI	8.15
No. of lifetime male sexual partners	CAPI	3.05
Received public assistance	ACASI	2.22
No. of times had sex in past 4 weeks	CAPI	1.37
Age of first sexual intercourse	CAPI	0.87
Blood tested for HIV	CAPI	0.65
Age of first menstrual period	CAPI	0.39
Highest grade completed	CAPI	0.04

Note. ACASI = audio computer-assisted self-interviewing; CAPI = computer-assisted personal interviewing.

As we noted, the item nonresponse rate is the highest for the total income question. This is quite consistent with prior work (Juster & Smith, 1997; Moore et al., 1999) and will come as no surprise to survey researchers. Questions about income are widely seen as very intrusive; in addition, some respondents may not know the household's income.

Response Quality

The best documented consequence of asking sensitive questions in surveys is systematic misreporting. Respondents consistently underreport some behaviors (the socially undesirable ones) and consistently overreport others (the desirable ones). This can introduce large biases into survey estimates.

Underreporting of socially undesirable behaviors appears to be quite common in surveys (see Tourangeau et al., 2000, chap. 9, for a review). Respondents seem to underreport the use of illicit drugs (Fendrich & Vaughn, 1994; L. D. Johnson & O'Malley, 1997), the consumption of alcohol (Duffy & Waterton, 1984; Lemmens, Tan, & Knibbe, 1992; Locander, Sudman & Bradburn, 1976), smoking (Bauman & Dent, 1982; Murray, O'Connell, Schmid, & Perry, 1987; Patrick et al., 1994), abortion (E. F. Jones & Forrest, 1992), bankruptcy (Locander et al., 1976), energy consumption (Warriner, McDougall, & Claxton, 1984), certain types of income (Moore et al., 1999), and criminal behavior (Wyner, 1980). They underreport racist attitudes as well (Krysan, 1998; see also Devine, 1989). By contrast, there is somewhat less evidence for overreporting of socially desirable behaviors in surveys. Still, overreporting has been found for reports about voting (Belli et al., 2001; Locander et al., 1976; Parry & Crossley, 1950; Traugott & Katosh, 1979), energy conservation (Fujii, Hennessy, & Mak, 1985), seat belt use (Stulginskas, Verreault, & Pless, 1985), having a library card (Locander et al., 1976; Parry & Crossley, 1950), church attendance (Presser & Stinson, 1998), and exercise (Tourangeau, Smith, & Rasinski, 1997). Many of the studies documenting underreporting or overreporting are based on comparisons of survey reports with outside records (e.g., Belli et al., 2001; E. F. Jones & Forrest, 1992; Locander et al., 1976; Moore et al., 1999; Parry & Crossley, 1950; Traugott & Katosh, 1979; Wyner, 1980) or physical assays (Bauman & Dent, 1982; Murray et al., 1987). From their review of the empirical findings, Tourangeau et al. (2000) concluded that response quality suffers as the topic becomes more sensitive and among those who have something to hide but that it can be improved by adopting certain design strategies. We review these strategies below.

Factors Affecting Reporting on Sensitive Topics

Survey researchers have investigated methods for mitigating the effects of question sensitivity on nonresponse and reporting error for more than 50 years. Their findings clearly indicate that several variables can reduce the effects of question sensitivity, decreasing item nonresponse and improving the accuracy of reporting. The key variables include the mode of administering the questions (especially whether or not an interviewer asks the questions), the data collection setting and whether other people are present as the respondent answers the questions, and the wording of the questions. Most of the studies we review below compare two or more different methods for eliciting sensitive information in a survey.

Many of these studies lack validation data and assume that whichever method yields more reports of the sensitive behavior is the more accurate method; survey researchers often refer to this as the "more is better" assumption.[2] Although this assumption is often plausible, it is still just an assumption.

Mode of Administration

Surveys use a variety of methods to collect data from respondents. Traditionally, three methods have dominated: face-to-face interviews (in which interviewers read the questions to the respondents and then record their answers on a paper questionnaire), telephone interviews (which also feature oral administration of the questions by interviewers), and mail surveys (in which respondents complete a paper questionnaire, and interviewers are not typically involved at all). This picture has changed radically over the last few decades as new methods of computer administration have become available and been widely adopted. Many national surveys that used to rely on face-to-face interviews with paper questionnaires have switched to computer-assisted personal interviewing (CAPI); in CAPI surveys, the questionnaire is no longer on paper but is a program on a laptop. Other face-to-face surveys have adopted a technique—audio computer-assisted self-interviewing (ACASI)—in which the respondents interact directly with the laptop. They read the questions on-screen and listen to recordings of the questions (typically, with earphones) and then enter their answers via the computer's keypad. A similar technique—interactive voice response (IVR)—is used in some telephone surveys. The respondents in an IVR survey are contacted by telephone (generally by a live interviewer) or dial into a toll-free number and then are connected to a system that administers a recording of the questions. They provide their answers by pressing a number on the keypad of the telephone or, increasingly, by saying aloud the number corresponding to their answer. In addition, some surveys now collect data over the Internet.

Although these different methods of data collection differ along a number of dimensions (for a thorough discussion, see chap. 5 in Groves, Fowler, et al., 2004), one key distinction among them is whether an interviewer administers the questions. Interviewer administration characterizes CAPI, traditional face-to-face interviews with paper questionnaires, and computer-assisted telephone interviews. By contrast, with traditional self-administered paper questionnaires (SAQs) and the newer computer-administered modes like ACASI, IVR, and Web surveys, respondents interact directly with the paper or electronic questionnaire. Studies going back nearly 40 years suggest that respondents are more willing to report sensitive information when the questions are self-administered than when they are administered by an interviewer (Hochstim, 1967). In the discussion that follows, we use the terms *computerized self-administration* and *computer administration of the questions* interchangeably. When the questionnaire is electronic but an interviewer reads the questions to the respondents and records their answers, we refer to it as *computer-assisted interviewing*.

---

[2] Of course, this assumption applies only to questions that are subject to underreporting. For questions about behaviors that are socially desirable and therefore overreported (such as voting), the opposite assumption is adopted—the method that elicits fewer reports is the better method.

Table 2

*Self-Administration and Reports of Illicit Drug Use: Ratios of Estimated Prevalence Under Self-
and Interviewer Administration, by Study, Drug, and Time Frame*

Study	Method of data collection	Drug	Month	Year	Lifetime
Aquilino (1994)	SAQ vs. FTF	Cocaine	1.00	1.50	1.14
		Marijuana	1.20	1.30	1.02
	SAQ vs. Telephone	Cocaine	1.00	1.00	1.32
		Marijuana	1.50	1.62	1.04
Aquilino & LoSciuto (1990)	SAQ vs. Telephone (Blacks)	Cocaine	1.67	1.22	1.21
		Marijuana	2.43	1.38	1.25
	SAQ vs. Telephone (Whites)	Cocaine	1.20	1.18	0.91
		Marijuana	1.00	1.04	1.00
Corkrey & Parkinson (2002)	IVR vs. CATI	Marijuana	—	1.58	1.26
Gfroerer & Hughes (1992)	SAQ vs. Telephone	Cocaine	—	2.21	1.43
		Marijuana	—	1.54	1.33
Schober et al. (1992)	SAQ vs. FTF	Cocaine	1.67	1.33	1.12
		Marijuana	1.34	1.20	1.01
Tourangeau & Smith (1996)	ACASI vs. FTF	Cocaine	1.74	2.84	1.81
		Marijuana	1.66	1.61	1.48
	CASI vs. CAPI	Cocaine	0.95	1.37	1.01
		Marijuana	1.19	0.99	1.29
Turner et al. (1992)	SAQ vs. FTF	Cocaine	2.46	1.58	1.05
		Marijuana	1.61	1.30	0.99

Note. Each study compares a method of self administration with a method of interviewer administration.
Dashes indicate that studies did not include questions about the illicit use of drugs during the prior month.
SAQ = self-administered paper questionnaires; FTF = face-to-face interviews with a paper questionnaire;
Telephone = paper questionnaires administered by a telephone interviewer; IVR = interactive voice response;
CATI = computer-assisted telephone interviews; ACASI = audio computer-assisted self-interviewing; CASI =
computer-assisted self-interviewing without the audio; CAPI = computer-assisted personal interviewing. From
"Reporting Issues in Surveys of Drug Use" by R. Tourangeau and T. Yan, in press, *Substance Use and Misuse,*
Table 3. Copyright 2005 by Taylor and Francis.

Table 2 (adapted from Tourangeau & Yan, in press) summarizes the results of several randomized field experiments that compare different methods for collecting data on illicit drug use.[3] The figures in the table are the ratios of the estimated prevalence of drug use under a self-administered mode to the estimated prevalence under some form of interviewer administration. For example, Corkrey and Parkinson (2002) compared IVR (computerized self-administration by telephone) with computer-assisted interviewer administration over the telephone and found that the estimated rate of marijuana use in the past year was 58% higher when IVR was used to administer the questionnaire. Almost without exception, the seven studies in Table 2 found that a higher proportion of respondents reported illicit drug use when the questions were self-administered than when they were administered by an interviewer. The median increase from self-administration is 30%. These results are quite consistent with the findings from a meta-analysis done by Richman, Kiesler, Weisband, and Drasgow (1999), who examined studies comparing computer-administered and interviewer-administered questionnaires. Their analysis focused on more specialized populations (such as psychiatric patients) than the studies in Table 2 and did not include any of the studies listed there. Still, they found a mean effect size of −.19, indicating greater reporting of psychiatric symptoms and socially undesirable behaviors when the computer administered the questions directly to the respondents:

> A key finding of our analysis was that computer instruments reduced social desirability distortion when these instruments were used as a substitute for face-to-face interviews, particularly when the interviews

were asking respondents to reveal highly sensitive personal behavior, such as whether they used illegal drugs or engaged in risky sexual practices. (Richman et al., 1999, p. 770)

An important feature of the various forms of self-administration is that the interviewer (if one is present at all) remains unaware of the respondent's answers. Some studies (e.g., Turner, Lessler, & Devore, 1992) have used a hybrid method in which an interviewer reads the questions aloud, but the respondent records the answers on a separate answer sheet; at the end of the interview, the respondent seals the answer sheet in an envelope. Again, the interviewer is never aware of how the respondent answered the questions. This method of self-administration seems just as effective as more conventional SAQs and presumably helps respondents with poor reading skills. ACASI offers similar advantages over paper questionnaires for respondents who have difficulty reading (O'Reilly, Hubbard, Lessler, Biemer, & Turner, 1994).

Increased levels of reporting under self-administration are apparent for other sensitive topics as well. Self-administration increases reporting of socially undesirable behaviors that, like illicit

---

[3] A general issue with mode comparisons is that the mode of data collection may affect not only reporting but also the level of unit or item nonresponse, making it difficult to determine whether the difference between modes reflects the impact of the method on nonresponse, reporting, or both. Most mode comparisons attempt to deal with this problem either by assigning cases to an experimental group after the respondents have agreed to participate or by controlling for any observed differences in the makeup of the different mode groups in the analysis.

drug use, are known to be underreported in surveys, such as abortions (Lessler & O'Reilly 1997; Mott, 1985) and smoking among teenagers (though the effects are weak, they are consistent across the major studies on teen smoking: see, e.g., Brittingham, Tourangeau, & Kay, 1998; Currivan, Nyman, Turner, & Biener, 2004; Moskowitz, 2004). Self-administration also increases respondents' reports of symptoms of psychological disorders, such as depression and anxiety (Epstein, Barker, & Kroutil, 2001; Newman et al., 2002). Problems in assessing psychopathology provided an early impetus to the study of social desirability bias (e.g., Jackson & Messick, 1961); self-administration appears to reduce such biases in reports about mental health symptoms (see, e.g., the meta-analysis by Richman et al., 1999). Self-administration can also reduce reports of socially desirable behaviors that are known to be overreported in surveys, such as attendance at religious services (Presser & Stinson, 1998).

Finally, self-administration seems to improve the quality of reports about sexual behaviors in surveys. As Smith (1992) has shown, men consistently report more opposite-sex sexual partners than do women, often by a substantial margin (e.g., ratios of more than two to one). This discrepancy clearly represents a problem: The *total* number of partners should be the same for the two sexes (because men and women are reporting the same pairings), and this implies that the *average* number of partners should be quite close as well (because the population sizes for the two sexes are nearly equal). The sex partner question seems to be sensitive in different directions for men and women, with men embarrassed to report too few sexual partners and women embarrassed to report too many. Tourangeau and Smith (1996) found that self-administration eliminated the gap between the reports of men and women, decreasing the average number of sexual partners reported by men and increasing the average number reported by women.

Differences Across Methods of Self-Administration

Given the variety of methods of self-administration currently used in surveys, the question arises whether the different methods differ among themselves. Once again, most of the large-scale survey experiments on this issue have examined reporting of illicit drug use. We did a meta-analysis of the relevant studies in an attempt to summarize quantitatively the effect of computerization versus paper self-administration across studies comparing the two.

Selection of studies. We searched for empirical reports of studies comparing computerized and paper self-administration, focusing on studies using random assignment of the subjects/survey respondents to one of the self-administered modes. We searched various databases available through the University of Maryland library (e.g., Ebsco, LexisNexis, PubMed) and supplemented this with online search engines (e.g., Google Scholar), using *self-administration*, *self-administered paper questionnaire*, *computer-assisted self-interviewing*, *interviewing mode*, *sensitive questions*, and *social desirability* as key words. We also evaluated the papers cited in the articles turned up through our search and searched the *Proceedings of the Survey Research Methods Section of the American Statistical Association*. These proceedings publish papers presented at the two major conferences for survey methodologists (the Joint Statistical Meetings and the annual conferences

of the American Association for Public Opinion Research) where survey methods studies are often presented.

We included studies in the meta-analysis if they randomly assigned survey respondents to one of the self-administered modes and if they compared paper and computerized modes of self-administration. We dropped two studies that compared different forms of computerized self-administration (e.g., comparing computer administered items with or without sound). We also dropped a few studies that did not include enough information to compute effect sizes. For example, two studies reported means but not standard deviations. Finally, we excluded three more studies because the statistics reported in the articles were not appropriate for our meta-analysis. For instance, Erdman, Klein, and Greist (1983) only reported agreement rates—that is, the percentages of respondents who gave the same answers under computerized and paper self-administration. A total of 14 studies met our inclusion criteria. Most of the papers we found were published, though not always in journals. We were not surprised by this; large-scale, realistic survey mode experiments tend to be very costly, and they are typically documented somewhere. We also used Duval and Tweedie's (2000) trim-and-fill procedure to test for the omission of unpublished studies; in none of our meta-analyses did we detect evidence of publication bias.

Of the 14 articles, 10 reported proportions by mode, and 4 reported mean responses across modes. The former tended to feature more or less typical survey items, whereas the latter tended to feature psychological scales. Table 3 presents a description of the studies included in the meta-analysis. Studies reporting proportions or percentages mostly asked about sensitive behaviors, such as using illicit drugs, drinking alcohol, smoking cigarettes, and so on, whereas studies reporting means tended to use various personality measures (such as self-deception scales, Marlow-Crowne Social Desirability scores, and so on). As a result, we conducted two meta-analyses. Both used log odds ratios as the effect size measure.

Analytic procedures. The first set of 10 studies contrasted the proportions of respondents reporting specific behaviors across different modes. Rather than using the difference in proportions as our effect size measures, we used the log odds ratios. Raw proportions are, in effect, unstandardized means, and this can inflate the apparent heterogeneity across studies (see the discussion in Lipsey & Wilson, 2001). The odds ratios (OR) were calculated as follows:

$$\text{OR} = p_{\text{computer}}(1 - p_{\text{paper}})/p_{\text{paper}}(1 - p_{\text{computer}}).$$

We then took the log of each odds ratio. For studies reporting means and standard errors, we converted the standardized mean effects to the equivalent log odds and used the latter as the effect size measure.

We carried out each of the meta-analyses we report here using the Comprehensive Meta-Analysis (CMA) package (Borenstein, Hedges, Higgins, & Rothstein, 2005). That program uses the simple average of the effect sizes from a single study (study *i*) to estimate the study effect size ($\hat{\bar{d}}_{i.}$ below); it then calculates the estimated overall mean effect size ($\hat{\bar{d}}_{..}$) as a weighted average of the study-level effect sizes:

Table 3
Descriptions and Mean Effect Sizes for Studies Included in the Meta-Analysis

Study	Sample size	Effect size	Sample type	Question type	Computerized mode
			Survey variables		
Beebe et al. (1998)	368	−0.28	Students at alternative schools	Behaviors (alcohol and drug use, victimization and crimes)	CASI
Chromy et al. (2002)	80,515	0.28	General population	Behaviors (drug use, alcohol, and cigarettes)	CASI
Evan & Miller (1969)	60	0.02	Undergraduates	Personality scales	Paper questionnaire, but computerized response entry
Kiesler & Sproull (1986)	50	−0.51	University students and employees	Values or attitudes	CASI
Knapp & Kirk (2003)	263	−0.33	Undergraduates	Behaviors (drug use, victimization, sexual orientation, been in jail, paid for sex)	Web, IVR
Lessler et al. (2000)	5,087	0.10	General population	Behaviors (drug use, alcohol, and cigarettes)	ACASI
Martin & Nagao (1989)	42	−0.22	Undergraduates	SAT and GPA overreporting	CASI
O'Reilly et al. (1994)	25	0.95	Volunteer students and general public	Behaviors (drug use, alcohol, and cigarettes)	ACASI, CASI
Turner et al. (1998)	1,729	0.74	Adolescent males	Sexual behaviors (e.g., sex with prostitutes, paid for sex, oral sex, male–male sex)	ACASI
Wright et al. (1998)	3,169	0.07	General population	Behaviors (drug use, alcohol, and cigarettes)	CASI
			Scale and personality measures		
Booth-Kewley et al. (1992)	164	−0.06	Male Navy recruits	Balanced Inventory of Desirable Responding Subscales (BIDR)	CASI
King & Miles (1995)	874	−0.09	Undergraduates	Personality measures (e.g., Mach V scale, Impression management)	CASI
Lautenschlager & Flaherty (1990)	162	0.43	Undergraduates	BIDR	CASI
Skinner & Allen (1983)	100	0.00	Adults with alcohol-related problems	Michigan Alcoholism Screening Test	CASI

Note. Each study compared a method of computerized self administration with a paper self-administered questionnaire (SAQ); a positive effect size indicates higher reporting under the computerized method of data collection. Effect sizes are log odds ratios. ACASI = audio computer-assisted self-interviewing; CASI = computer-assisted self-interviewing without the audio; IVR = interactive voice response.

$$\hat{d}_{..} = \frac{\sum\limits^{l} w_i \hat{d}_{i.}}{\sum\limits^{l} w_i} = \frac{\sum\limits^{l} w_i \sum\limits^{k_i} d_{ij}/k_i}{\sum\limits^{l} w_i}$$

$$w_i = \frac{1}{\sum\limits_{k_i} \hat{\sigma}_{ij}^2 / k_i}$$

$$= \frac{k_i}{\sum\limits_{k_i} \hat{\sigma}_{ij}^2}. \qquad (1)$$

The study-level weight (w_i) is the inverse of the mean of the variances ($\hat{\sigma}_{ij}^2$) of the k_i effect size estimates from that study. There are a couple of potential problems with this approach. First, the program computes the variance of each effect size estimate as though the estimates were derived from simple

random samples. This assumption is violated by most surveys and many of the methodological studies we examined, which feature stratified, clustered, unequal probability samples. As a result, the estimated variances ($\hat{\sigma}_{ij}^2$) tend to be biased. Second, the standard error of the final overall estimate ($\hat{d}_{..}$) depends primarily on the variances of the individual estimates (see, e.g., Lipsey & Wilson, 2001, p. 114). Because these variances are biased, the overall standard errors are biased as well.

To assess the robustness of our findings, we also computed the overall mean effect size estimate using a somewhat different weighting scheme:

$$\hat{d}'_{..} = \frac{\sum\limits^{l} \sum\limits^{k_i} w_{ij} \hat{d}_{ij}/k_i}{\sum\limits^{l} \sum\limits^{k_i} w_{ij}/k_i}$$

$$= \frac{\sum\limits^{l} 1/k_i \sum\limits^{k_i} w_{ij}\hat{d}_{ij}}{\sum\limits^{l} 1/k_i \sum\limits^{k_i} w_{ij}} = \frac{\sum\limits^{l} 1/k_i \sum\limits^{k_i} w_{ij}\hat{d}_{ij}}{\sum\limits^{l} \bar{w}_{i.}}$$

$$w_{ij} = \frac{1}{\hat{\sigma}_{ij}^2}. \qquad (2)$$

This weighting strategy gives less weight to highly variable effect sizes than the approach in Equation 1, though it still assumes simple random sampling in calculating the variances of the individual estimates. We therefore also used a second program—SAS's PROC SURVEYMEANS—to calculate the standard error of the overall effect size estimate $(\hat{\bar{d}}')$ given in Equation 2. This program uses the variation in the (weighted) mean effect sizes across studies to calculate a standard error for the overall estimate, without making any assumptions about the variability of the individual estimates; within the sampling literature, PROC SURVEYMEANS provides a "design-based" estimate of the standard error of $\hat{\bar{d}}'$ (see, e.g., Wolter, 1985). For the most part, the two methods for carrying the meta-analyses yield similar conclusions and we note the one instance where they diverge.

Results. Overall, the mean effect size for computerization in the studies examining reports about specific behaviors is 0.08 (with a standard error of 0.07); across the 10 studies, there is a nonsignificant tendency for computerized self-administration to elicit more socially undesirable responses than paper self-administration. The 4 largest survey experiments all found positive effects for computerized self-administration, with mean effect sizes ranging from 0.07 (Wright, Aquilino, & Supple, 1998) to 0.28 (Chromy, Davis, Packer, & Gfroerer, 2002). For the 4 studies examining personality scales (including measures of socially desirable responding), computerization had no discernible effect relative to paper self-administration (the mean effect size was −0.02, with a standard error of 0.10). Neither set of studies exhibits significant heterogeneity (Q = 6.23 for the 10 studies in the top panel; Q = 2.80 for the 4 studies in the bottom panel).[4]

The meta-analysis by Richman et al. (1999) examined a different set of studies from those in Table 3. They deliberately excluded papers examining various forms of computerized self-administration such as ACASI, IVR, and Web surveys and focused on studies that administered standard psychological batteries, such as the Minnesota Multiphasic Personality Inventory, rather than sensitive survey items. They nonetheless arrived at similar conclusions about the impact of computerization. They reported a mean effect size of 0.05 for the difference between computer and paper self-administration, not significantly different from zero, with the overall trend slightly in favor of self-administration on paper. In addition, they found that the effect of computer administration depended on other features of the data collection situation, such as whether the respondents were promised anonymity. When the data were not anonymous, computer administration decreased the reporting of sensitive information relative to paper self-administered questionnaires; storing identifiable answers in an electronic data base may evoke the threat of "big brother" and discourage reporting of sensitive information (see Rosenfeld, Booth-Kewley, Edwards, & Thomas, 1996). Generally, surveys promise confidentiality rather than complete anonymity, but the

benefits may be similar. In a related result, Moon (1998) also found that respondents seem to be more skittish about reporting socially undesirable information when they believe their answers are being recorded by a distant computer rather than a stand-alone computer directly in front of them.

Interviewer Presence, Third Party Presence, and Interview Setting

The findings on mode differences in reporting of sensitive information clearly point a finger at the interviewer as a contributor to misreporting. It is not that the interviewer does anything wrong. What seems to make a difference is whether the respondent has to report his or her answers to another person. When an interviewer is physically present but unaware of what the respondent is reporting (as in the studies with ACASI), the interviewer's mere presence does not seem to have much effect on the answers (although see Hughes, Chromy, Giacolletti, & Odom, 2002, for an apparent exception).

What if the interviewer is aware of the answers but not physically present, as in a telephone interview? The findings on this issue are not completely clear, but taken together, they indicate that the interviewer's physical presence is not the important factor. For example, some of the studies suggest that telephone interviews are less effective than face-to-face interviews in eliciting sensitive information (e.g., Aquilino, 1994; Groves & Kahn, 1979; T. Johnson, Hougland, & Clayton, 1989), but a few studies have found the opposite (e.g., Henson, Cannell, & Roth, 1978; Hochstim, 1967; Sykes & Collins, 1988), and still others have found no differences (e.g., Mangione, Hingson, & Barrett, 1982). On the whole, the weight of the evidence suggests that the telephone interviews yield less candid reporting of sensitive information (see the meta-analysis by de Leeuw & van der Zouwen, 1988). We found only one new study (Holbrook et al., 2003) on the differences between telephone and face-to-face interviewing that was not covered in de Leeuw and van der Zouwen's (1988) meta-analysis and did not undertake a new meta-analysis on this issue. Holbrook et al. come to the same conclusion as did de Leeuw and van der Zouwen—that socially desirability bias is worse in telephone interviews than in face-to-face interviews.

It is virtually an article of faith among survey researchers that no one but the interviewer and respondent should be present during the administration of the questions, but survey interviews are often conducted in less than complete privacy. According to a study by Gfroerer (1985), only about 60% of the interviews in the 1979 and 1982 National Surveys on Drug Abuse were rated by the interviewers as having been carried out in complete privacy. Silver, Abramson, and Anderson (1986) reported similar figures for the American National Election Studies; their analyses covered the period from 1966 to 1982 and found that roughly half the inter-

---

[4] The approach that uses the weight described in Equation 2 and PROC SURVEYMEANS yielded a different conclusion from the standard approach. The alternative approach produced a mean effect size estimate of .08 (consistent with the estimate from the CMA program), but a much smaller standard error estimate (0.013). These results indicate a significant increase in the reporting of sensitive behaviors for computerized relative to paper self-administration.

views were done with someone else present besides the respondent and interviewer.

Aquilino, Wright, and Supple (2000) argued that the impact of the presence of other people is likely to depend on whether the bystander already knows the information the respondent has been asked to provide and whether the respondent fears any repercussions from revealing it to the bystander. If the bystander already knows the relevant facts or is unlikely to care about them, the respondent will not be inhibited from reporting them. In their study, Aquilino et al. found that the presence of parents led to reduced reporting of alcohol consumption and marijuana use; the presence of a sibling or child had fewer effects on reporting; and the presence of a spouse or partner had no apparent effect at all. As they had predicted, computer-assisted self-interviewing without audio (CASI) seemed to diminish the impact of bystanders relative to an SAQ: Parental presence had no effect on reports about drug use when interview was conducted with CASI but reduced reported drug use when the data were collected in a paper SAQ. Presumably, the respondents were less worried that their parents would learn their answers when the information disappeared into the computer than when it was recorded on a paper questionnaire. In order to get a quantitative estimate of the overall effects of bystanders on reporting, we carried out a second meta-analysis.

Selection of studies and analytic procedures. We searched for published and unpublished empirical studies that reported the effect of bystanders on reporting, using the same databases as in the meta-analysis on the effects of computerized self-administration on reporting. This time we used the keywords *bystander*, *third-party*, *interview privacy*, *spouse presence*, *privacy*, *presence*, *survey*, *sensitivity*, *third-party presence during interview*, *presence of another during survey interview*, and *effects of bystanders in survey interview*. We focused on research articles

that specified the type of bystanders (e.g., parent versus spouse), as that is a key moderator variable; we excluded papers that examined the effects of group setting and the effects of anonymity on reporting from the meta-analysis, but we discuss them below. We again used Duval and Tweedie's (2000) trim-and-fill procedure to test for publication bias in the set of studies we examined and found no evidence for it.

Our search turned up nine articles that satisfied our inclusion criteria. We carried out separate meta-analyses for different types of bystanders and focussed on the effects of spouse and parental presence, as there are a relatively large number of studies on these types of bystanders. We used log odds ratios as the effect size measure, weighted each effect size estimate (using the approach in Equation 1 above), and used CMA to carry out the analyses. (We also used the weight in Equation 2 above and PROC SURVEYMEANS and reached the same conclusions as those reported below.)

Results. As shown in Table 4, the presence of a spouse does not have a significant overall impact on survey responses, but the studies examining this issue show significant heterogeneity ($Q = 10.66$, $p < .05$). One of the studies (Aquilino, 1993) has a large sample size and asks questions directly relevant to the spouse. When we dropped this study, the effect of spouse presence remained nonsignificant (under the random effects model and under the PROC SURVEYMEANS approach), but tended to be positive (that is, on average, the presence of a spouse increased reports of sensitive information); the heterogeneity across studies is no longer significant. Parental presence, by contrast, seems to reduce socially undesirable responses (see Table 5); the effect of parental presence is highly significant ($p < .001$). The presence of children does not seem to affect survey responses, but the number of studies

Table 4

Spousal Presence Characteristics and Effect Sizes for Studies Included in the Meta-Analysis of the Effects of Bystanders

Study	Sample size	Effect size	Sample type	Question type
Aquilino (1993)	6,593	−0.02	General population	Attitudes on cohabitation and separation
Aquilino (1997)	1,118	0.44	General population	Drug use
Aquilino et al. (2000)	1,026	0.28	General population	Alcohol and drug use
Silver et al. (1986)	557	−0.01	General population	Voting
Taietz (1962)	122	−1.36	General population (in the Netherlands)	Attitudes

	Effect size			
Study and effect	*M*	*SE*	*Z*	Homogeneity test
Including all studies				
Fixed	0.12	0.08	1.42	$Q = 10.66$, $p = .03$
Random	0.10	0.16	0.64	
Excluding Aquilino (1993)				
Fixed	0.27	0.12	2.28	$Q = 7.43$, $p = .06$
Random	0.13	0.22	0.59	

Note. Each study compared responses when a bystander was present with those obtained when no bystander was present; a positive effect size indicates higher reporting when a particular type of bystander was present. Effect sizes are log odds ratios.

Table 5

Parental Presence Characteristics and Effect Sizes for Studies Included in the Meta-Analysis of the Effects of Bystanders

Study	Sample size	Effect size	Sample type	Question type
Aquilino (1997)	521	−0.55	General population	Drug use
Aquilino et al. (2000)	1,026	−0.77	General population	Alcohol and drug use
Gfroerer (1985)	1,207	−0.60	Youth (12–17 years old)	Drug and alcohol use
Moskowitz (2004)	2,011	−0.25	Adolescents	Smoking
Schutz & Chilcoat (1994)	1,287	−0.70	Adolescents	Behaviors (drug, alcohol, and tobacco)

	Effect size			
Effect	*M*	*SE*	*Z*	Homogeneity test
Fixed	−0.38	0.08	−5.01	$Q = 7.02, p = .13$
Random	−0.50	0.13	−3.83	

Note. Each study compared responses when a bystander was present with those obtained when no bystander was present; a positive effect size indicates higher reporting when a particular type of bystander was present. Effect sizes are log odds ratios.

examining this issue is too small to draw any firm conclusions (see Table 6).

Other studies have examined the effects of collecting data in group settings such as schools. For example, Beebe et al. (1998) found that computer administration elicited somewhat fewer reports of sensitive behaviors (like drinking and illicit drug use) from high school students than paper SAQs did. In their study, the questions were administered in groups at school, and the computers were connected to the school network. In general, though, schools seem to be a better place for collecting sensitive data from high school students than their homes. Three national surveys monitor trends in drug use among high students. Two of the them—Monitoring the Future (MTF) and the Youth Risk Behavior Survey (YRBS)—collect the data in schools via paper SAQs; the third—the National Household Survey on Drug Abuse—collects them in the respondent's home using ACASI. The two surveys done in school setting have yielded estimated levels of smoking and illicit drug use that are considerably higher than those obtained from the household survey (Fendrich & Johnson, 2001; Fowler & Stringfellow, 2001; Harrison, 2001; Sudman, 2001). Table 7 (derived from Fendrich & Johnson, 2001) shows some typical findings, comparing estimates of lifetime and recent use of cigarettes, marijuana, and cocaine. The differences across the surveys are highly significant but may reflect other differences across the surveys than the data collection setting. Still, a recent experimental comparison (Brener et al., 2006) provides evidence that the setting is the key variable. It is possible that moving the survey away from the presence of parents reduces underreporting of risky behaviors; it is also possible that the presence of peers in school settings leads to overreporting of such behaviors.[5]

All of the findings we have discussed so far on the impact of the presence of other people during data collection, including those in Tables 4–6, are from nonexperimental studies. Three additional studies attempted to vary the level of the privacy of the data collection setting experimentally. Mosher and Duffer (1994) and Tourangeau, Rasinski, Jobe, Smith, and Pratt (1997) randomly assigned respondents to be interviewed either at home or in a neutral setting outside the home, with the latter guaranteeing privacy (at least from other family members). Neither study found clear effects on reporting for the site of the interview. Finally, Couper, Singer, and Tourangeau (2003) manipulated whether a bystander (a confederate posing as a computer technician) came into the room where the respondent was completing a sensitive questionnaire. Respondents noticed the intrusion and rated this condition significantly less private than the condition in which no interruption took place, but the confederate's presence had no discernible effect on the reporting of sensitive information. These findings are more or less consistent with the model of bystander effects proposed by Aquilino et al. (2000) in that respondents may not believe they will experience any negative consequences from revealing their responses to a stranger (in this case, the confederate). In addition, the respondents in the study by Couper et al. (2003) answered the questions via some form of CASI (either with sound and text, sound alone, or text alone), which may have eliminated any impact that the presence of the confederate might otherwise have had.

The Concept of Media Presence

The findings on self-administration versus interviewer administration and on the impact of the presence of other people suggest the key issue is not the physical presence of an interviewer (or a bystander). An interviewer does not have to be physically present to inhibit reporting (as in a telephone interview), and interviewers

---

[5] It is not clear why the two school-based surveys differ from each other, with YRBS producing higher estimates than MTF on all the behaviors in Table 7. As Fendrich and Johnson (2001) noted, the MTF asks for the respondents' name on the form, whereas the YRBS does not; this may account for the higher levels of reporting in the YRBS. This would be consistent with previous findings about the impact of anonymity on reporting of socially undesirable information (e.g., Richman et al., 1999).

Table 6
Children Presence Characteristics and Effect Sizes for Studies Included in the Meta-Analysis of the Effects of Bystanders

Study	Sample size	Effect size	Sample type	Question type
Aquilino et al. (2000)	1,024	−1.04	General population	Alcohol and drug use
Taietz (1962)	122	0.63	General population (in the Netherlands)	

Effect	Effect size		Z	Homogeneity test
	M	SE		
Fixed	0.21	0.43	0.49	Q = 2.83, p = .09
Random	−0.06	0.83	−0.07	

Note. Each study compared responses when a bystander was present with those obtained when no bystander was present; a positive effect size indicates higher reporting when a particular type of bystander was present. Effect sizes are log odds ratios.

or bystanders who are physically present but unaware of the respondent's answers seem to have diminished impact on what the respondent reports (as in the various methods of self-administration). What seems to matter is the threat that someone with whom the respondent will continue to interact (an interviewer or a bystander) will learn something embarrassing about the respondent or will learn something that could lead them to punish the respondent in some way.

A variety of new technologies attempt to create an illusion of presence (such as virtual reality devices and video conferencing). As a result, a number of computer scientists and media researchers have begun to explore the concept of media presence. In a review of these efforts, Lombard and Ditton (1997) argued that media presence has been conceptualized in six different ways. A sense of presence is created when (a) the medium is seen as socially rich (e.g., warm or intimate), (b) it conveys realistic representations of events or people, (c) it creates a sense that the user or a distant object or event has been transported (so that one is really there in

another place or that a distant object is really with the user), (d) it immerses the user in sensory input, (e) it makes mediated or artificial entities seem like real social actors (such as actors on television or the cyber pets popular in Japan), or (f) the medium itself comes to seem like a social actor. These conceptualizations, according to Lombard and Ditton, reflect two underlying views of media presence. A medium can create a sense of presence by becoming transparent to the user, so that the user loses awareness of any mediation; or, at the other extreme, the medium can become a social entity itself (as when people argue with their computers). A medium like the telephone is apparently transparent enough to convey a sense of a live interviewer's presence; by contrast, a mail questionnaire conveys little sense of the presence of the researchers and is hardly likely to be seen as a social actor itself.

As surveys switch to modes of data collection that require respondents to interact with ever more sophisticated computer interfaces, the issue arises whether the survey questionnaire or the computer might come to be treated as social actors. If so, the gains in reporting sensitive information from computer administration of the questions may be lost. Studies by Nass, Moon, and Green (1997) and by Sproull, Subramani, Kiesler, Walker, and Waters (1996) demonstrate that seemingly incidental features of the computer interface, such as the gender of the voice it uses (Nass et al., 1997), can evoke reactions similar to those produced by a person (e.g., gender stereotyping). On the basis of the results of a series of experiments that varied features of the interface in computer tutoring and other tasks, Nass and his colleagues (Fogg & Nass, 1997; Nass et al., 1997; Nass, Fogg, & Moon, 1996; Nass, Moon, & Carney, 1999) argued that people treat computers as social actors rather than as inanimate tools. For example, Nass et al. (1997) varied the sex of the recorded (human) voice through which a computer delivered tutoring instructions to the users; the authors argue that the sex of the computer's voice evoked gender stereotypes toward the computerized tutor. Walker, Sproull, and Subramani (1994) reported similar evidence that features of the interface can evoke the presence of a social entity. Their study varied whether the computer presented questions using a text display or via one of two talking faces. The talking face displays led users to spend more time interacting with the program and decreased the

Table 7
Estimated Rates of Lifetime and Past Month Substance Use among 10th Graders, by Survey

Substance and use	MTF	YRBS	NHSDA
Cigarettes			
Lifetime	60.2%	70.0%	46.9%
Past month	29.8%	35.3%	23.4%
Marijuana			
Lifetime	42.3%	46.0%	25.7%
Past month	20.5%	25.0%	12.5%
Cocaine			
Lifetime	7.1%	7.6%	4.3%
Past month	2.0%	2.6%	1.0%

Note. Monitoring the Future (MTF) and the Youth Risk Behavior Survey (YRBS) are done in school; the National Household Survey of Drug Abuse (NHSDA) is done in the respondent's home. Data are from "Examining Prevalence Differences in Three National Surveys of Youth: Impact of Consent Procedures, Mode, and Editing Rules," by M. Fendrich and T. P. Johnson, 2001, *Journal of Drug Issues, 31,* 617. Copyright 2001 by Florida State University.

number of errors they made. The users also seemed to prefer the more expressive face to a relatively impassive one (see also Sproull et al., 1996, for related results). These studies suggest that the computer can exhibit "presence" in the second of the two broad senses distinguished by Lombard and Ditton (1997).

It is not clear that the results of these studies apply to survey interfaces. Tourangeau, Couper, and Steiger (2003) reported a series of experiments embedded in Web and IVR surveys that varied features of the electronic questionnaire, but they found little evidence that respondents treated the interface or the computer as social actors. For example, their first Web experiment presented images of a male or female researcher at several points throughout the questionnaire, contrasting both of these with a more neutral image (the study logo). This experiment also varied whether the questionnaire addressed the respondent by name, used the first person in introductions and transitional phrases (e.g., "Thanks, Roger. Now I'd like to ask you a few questions about the roles of men and women."), and echoed the respondents' answers ("According to your responses, you exercise once daily."). The control version used relatively impersonal language ("The next series of questions is about the roles of men and women.") and gave feedback that was not tailored to the respondent ("Thank you for this information."). In another experiment, Tourangeau et al. varied the gender of the voice that administered the questions in an IVR survey. All three of their experiments used questionnaires that included questions about sexual behavior and illicit drug use as well as questions on gender-related attitudes. Across their three studies, Tourangeau et al. found little evidence that the features of the interface affected reporting about sensitive behaviors or scores on the BIDR Impression Management scale, though they did find a small but consistent effect on responses to the gender attitude items. Like live female interviewers, "female" interfaces (e.g., a recording of a female voice in an IVR survey) elicited more pro-feminist responses on these items than "male" interfaces did; Kane and Macaulay (1993) found similar gender-of-interviewer effects for these items with actual interviewers. Couper, Singer, and Tourangeau (2004) reported further evidence that neither the gender of the voice nor whether it was a synthesized or a recorded human voice affected responses to sensitive questions in an IVR survey.

Indirect Methods for Eliciting Sensitive Information

Apart from self-administration and a private setting, several other survey design features influence reports on sensitive topics. These include the various indirect strategies for eliciting the information (such as the randomized response technique [RRT]) and the bogus pipeline.

The randomized response technique was introduced by Warner (1965) as a method for improving the accuracy of estimates about sensitive behaviors. Here is how Warner described the procedure:

> Suppose that every person in the population belongs to either Group A or Group B . . . Before the interviews, each interviewer is furnished with an identical spinner with a face marked so that the spinner points to the letter A with probability p and to the letter B with probability $(1 - p)$. Then, in each interview, the interviewee is asked to spin the spinner unobserved by the interviewer and report only whether the spinner points to the letter representing the group to which the interviewee belongs. That is, the interviewee is required only to say

> yes or no according to whether or not the spinner points to the correct group. (Warner, 1965, p. 64)

For example, the respondent might receive one of two statements about a controversial issue (A: I am for legalized abortion on demand; B: I am against legalized abortion on demand) with fixed probabilities. The respondent reports whether he or she agrees with the statement selected by the spinner (or some other randomizing device) without revealing what the statement is. Warner did not actually use the technique but worked out the mathematics for deriving an estimated proportion (and its variance) from data obtained this way. The essential feature of the technique is that the interviewer is unaware of what question the respondent is answering. There are more efficient (that is, lower variance) variations on Warner's original theme. In one—the unrelated question method (Greenberg, Abul-Ela, Simmons, & Horvitz, 1969)—the respondent answers either the sensitive question or an unrelated innocuous question with a known probability of a "yes" answer ("Were you born in April?"). In the other common variation—the forced alternative method—the randomizing device determines whether the respondent is supposed to give a "yes" answer, a "no" answer, or an answer to the sensitive question. According to a meta-analysis of studies using RRT (Lensvelt-Mulders, Hox, van der Heijden, & Maas, 2005), the most frequently used randomizing devices are dice and coins. Table 8 below shows the formulas for deriving estimates from data obtained via the different RRT procedures and also gives the formulas for estimating the variance of those estimates.

The meta-analysis by Lensvelt-Mulders et al. (2005) examined two types of studies that evaluated the effectiveness of RRT relative to other methods for collecting the same information, such as standard face-to-face interviewing. In one type, the researchers had validation data and could determine the respondent's actual status on the variable in question. In the other type, there were no true scores available, so the studies compare the proportions of respondents admitting to some socially undesirable behavior or attitude under RRT and under some other method, such as a direct question in a face-to-face interview. Both types of studies indicate that RRT reduces underreporting compared to face-to-face interviews. For example, in the studies with validation data, RRT produced a mean level of underreporting of 38% versus 42% for face-to-face interviews. The studies without validation data lead to similar conclusions. Although the meta-analysis also examined other modes of data collection (such as telephone interviews and paper SAQs), these comparisons are based on one or two studies or did not yield significant differences in the meta-analysis.

These gains from RRT come at a cost. As Table 8 makes clear, the variance of the estimate of the prevalence of the sensitive behavior is increased. For example, with the unrelated question method, even if 70% of the respondents get the sensitive item, this is equivalent to reducing the sample size by about half (that is, by a factor of .70^2). RRT also makes it more difficult to estimate the relation between the sensitive behavior and other characteristics of the respondents.

The virtue of RRT is that the respondent's answer does not reveal anything definite to the interviewer. Several other methods for eliciting sensitive information share this feature. One such method is called the *item count technique* (Droitcour et al., 1991); the same method is also called the *unmatched count technique*

Table 8
Estimators and Variance Formulas for Randomized Response and Item Count Methods

Method	Estimator	Variance
	Randomized response technique	
Warner's method	$\hat{p}_w = \dfrac{\hat{\lambda} - 1 - \pi}{2\pi - 1}$ π = probability that the respondent receives statement A $\hat{\lambda}$ = observed percent answering "yes"	$Var(\hat{p}_w) = \dfrac{\hat{\lambda}(1 - \hat{\lambda})}{n} + \dfrac{\pi(1 - \pi)}{n(2\pi - 1)^2}$
Unrelated question	$\hat{p}_u = \dfrac{\hat{\lambda} - p_2(1 - \pi)}{\pi}$ π = probability that the respondent gets the sensitive question p_2 = known probability of "yes" answer on unrelated question $\hat{\lambda}$ = observed percent answering "yes"	$Var(\hat{p}_u) = \dfrac{\hat{\lambda}(1 - \hat{\lambda})}{n\pi^2}$
Forced choice	$\hat{p}_{FC} = \dfrac{\hat{\lambda} - \pi_1}{\pi_2}$ π_1 = probability that the respondent is told to say "yes" π_2 = probability that respondent is told to answer sensitive question $\hat{\lambda}$ = observed percent answering "yes"	$Var(\hat{p}_{FC}) = \dfrac{\hat{\lambda}(1 - \hat{\lambda})}{n\pi_2^2}$
	Item count	
One list	$\hat{p}_{1L} = \bar{x}_{k+1} - \bar{x}_k$	$Var(\hat{p}_{1L}) = Var(\bar{x}_{k+1}) + Var(\bar{x}_k)$
Two lists	$\hat{p}_{2L} = (\hat{p}_1 + \hat{p}_2)/2$ $\hat{p}_1 = \bar{x}_{1,k+1} - \bar{x}_{1,k}$ $\hat{p}_2 = \bar{x}_{2,k+1} - \bar{x}_{2,k}$ The list with the subscript $k+1$ includes the sensitive item; the list with the subscript k omits it	$Var(\hat{p}_{2L}) = \dfrac{Var(\hat{p}_1) + Var(\hat{p}_2) + 2\rho_{12}\sqrt{Var(\hat{p}_1)Var(\hat{p}_2)}}{4}$

(Dalton, Daily, & Wimbush, 1997). This procedure asks one group of respondents to report how many behaviors they have done on a list that includes the sensitive behavior; a second group reports how many behaviors they have done on the same list of behaviors, omitting the sensitive behavior. For example, the question might ask, "How many of the following have you done since January 1: Bought a new car, traveled to England, donated blood, gotten a speeding ticket, and visited a shopping mall?" (This example is taken from Droitcour et al., 1991.) A second group of respondents gets the same list of items except for the one about the speeding ticket. Typically, the interviewer would display the list to the respondents on a show card. The difference in the mean number of items reported by the two groups is the estimated proportion reporting the sensitive behavior. To minimize the variance of this estimate, the list should include only one sensitive item, and the other innocuous items should be low-variance items— that is, they should be behaviors that nearly everyone has done or that nearly everyone has not done (see Droitcour et al., 1991, for a discussion of the estimation issues; the relevant formulas are given in Table 8).

A variation on the basic procedure uses two pairs of lists (see Biemer & Brown, 2005, for an example). Respondents are again randomly assigned to one of two groups. Both groups get two lists of items. For one group of respondents, the sensitive item is included in the first list; for the other, it is included in the second list. This method produces two estimates of the proportion of respondents reporting the sensitive behavior. One is based on the difference between the two groups on the mean number of items reported from the first list. The other is based on the mean difference in the number of items reported from the second list. The two estimates are averaged to produce the final estimate.

Meta-analysis methods. Only a few published studies have examined the effectiveness of the item count method, and they seem to give inconsistent results. We therefore conducted a meta-analysis examining the effectiveness of this technique. We searched the same databases as in the previous two meta-analyses, using *item count, unmatched count, two-list, sensitive questions,* and *survey* (as well as combinations of these) as our key words. We restricted ourselves to papers that (a) contrasted the item count technique with more conventional methods of asking sensitive questions and (b) reported quantitative estimates under both methods. Seven papers met these inclusion criteria. Once again, we used log odds ratios as the effect size measure, weighted each effect size (to reflect its variance and the number of estimates from each study), and used the CMA software to compute the estimates. (In addition, we carried out the meta-analyses using our alternative approach, which yielded similar results.)

Results. Table 9 displays the key results from the meta-analysis. They indicate that the use of item count techniques generally elicits more reports of socially undesirable behaviors

Table 9
Characteristics and Effect Sizes for Studies Included in the Meta-Analysis of the Item Count Technique

Study	Sample size	Effect size	Sample sype	Question type
Ahart & Sackett (2004)	318	−0.16	Undergraduates	Dishonest behaviors (e.g., calling in sick when not ill)
Dalton et al. (1994)	240	1.58	Professional auctioneers	Unethical behaviors (e.g., conspiracy nondisclosure)
Droitcour et al. (1991)	1,449	−1.51	General population (in Texas)	HIV risk behaviors
Labrie & Earleywine (2000)	346	0.39	Undergraduates	Sexual risk behaviors and alcohol
Rayburn et al. (2003a)	287	1.51	Undergraduates	Hate crime victimizations
Rayburn et al. (2003b)	466	1.53	Undergraduates	Anti-gay hate crimes
Wimbush & Dalton (1997)	563	0.97	Past and present employees	Employee theft

Study and effect	Point estimates	SE	z	Homogeneity test
All seven studies				
Fixed	0.26	0.14	1.90	Q = 35.45, $p < .01$
Random	0.45	0.38	1.19	
Four studies with undergraduate samples				
Fixed	0.19	0.17	1.07	Q = 6.03, $p = 0.11$
Random	0.33	0.31	1.08	
Other three studies				
Fixed	0.40	0.23	1.74	Q = 28.85, $p < .01$
Random	0.35	0.90	0.38	

than direct questions; but the overall effect is not significant, and there is significant variation across studies (Q = 35.4, $p < .01$). The studies using undergraduate samples tended to yield positive results (that is, increased reporting under the item count technique), but the one general population survey (reported by Droitcour et al., 1991) received clearly negative results. When we partition the studies into those using undergraduates and those using other types of sample, neither group of studies shows a significant overall effect, and the three studies using nonstudent samples continue to exhibit significant heterogeneity (Q = 28.8, $p < .01$). Again, Duval and Tweedie's (2000) trim-and-fill procedure produced no evidence of publication bias in the set of studies we examined.

The three-card method. A final indirect method for collecting sensitive information is the three-card method (Droitcour, Larson, & Scheuren, 2001). This method subdivides the sample into three groups, each of which receives a different version of a sensitive question. The response options offered to the three groups combine the possible answers in different ways, and no respondent is directly asked about his or her membership in the sensitive category. The estimate of the proportion in the sensitive category is obtained by subtraction (as in the item count method). For example, the study by Droitcour et al. (2001) asked respondents about their immigration status; here the sensitive category is "illegal alien." None of the items asked respondents whether they were illegal aliens. Instead, one group got an item that asked whether they had a green card or some other status (including U.S. citizenship); a second group got an item asking them whether they were U.S. citizens or had some other status; the final group got an item asking whether they had a valid visa (student, refugee, tourist, etc.). The estimated proportion of illegal aliens is 1 minus the sum

of the proportions with green cards (based on the answers from Group 1), the proportion who are U.S. citizens (based on the answers from Group 2), and the proportion with valid visas (based on the answers from Group 3). We do not know of any studies that compare the three-card method to direct questions.

Other indirect question strategies. Survey textbooks (e.g., Sudman & Bradburn, 1982) sometimes recommend two other indirect question strategies for sensitive items. One is an analogue to RRT for items requiring a numerical response. Under this method (the "additive constants" or "aggregated response" technique), respondents are instructed to add a randomly generated constant to their answer (see Droitcour et al., 1991, and Lee, 1993, for descriptions). If the distribution of the additive constants is known, then an unbiased estimate of the quantity of interest can be extracted from the answers. The other method, the "nominative method," asks respondents to report on the sensitive behaviors of their friends ("How many of your friends used heroin in the last year?"). An estimate of the prevalence of heroin use is derived via standard multiplicity procedures (e.g., Sirken, 1970). We were unable to locate any studies evaluating the effectiveness of either method for collecting sensitive information.

The Bogus Pipeline

Self-administration, RRT, and the other indirect methods of questioning respondents all prevent the interviewer (and, in some cases, the researchers) from learning whatever sensitive information an individual respondent might have reported. In fact, RRT and the item count method produce only aggregate estimates rather than individual scores. The bogus pipeline works on a different principle; with this method, the respondent believes that the inter-

viewer will learn the respondent's true status on the variable in question regardless of what he or she reports (Clark & Tifft, 1966; see E. E. Jones & Sigall, 1971, for an early review of studies using the bogus pipeline, and Roese and Jamieson, 1993, for a more recent one). Researchers have used a variety of means to convince the respondents that they can detect false reports, ranging from polygraph-like devices (e.g., Tourangeau, Smith, & Rasinski, 1997) to biological assays that can in fact detect false reports (such as analyses of breath or saliva samples that can detect recent smoking; see Bauman & Dent, 1982, for an example).[6] The bogus pipeline presumably increases the respondent's motivation to tell the truth—misreporting will only add the embarrassment of being caught out in a lie to the embarrassment of being exposed as an illicit drug user, smoker, and so forth.

Roese and Jamieson's (1993) meta-analysis focused mainly on sensitive attitudinal reports (e.g., questions about racial prejudice) and found that the bogus pipeline significantly increases respondents' reports of socially undesirable attitudes (Roese & Jamieson, 1993). Several studies have also examined reports of sensitive behaviors (such as smoking, alcohol consumption, and illicit drug use). For example, Bauman and Dent (1982) found that the bogus pipeline increased accuracy in reports by teens about smoking. They tested breath samples to determine whether the respondents had smoked recently; in their study, the "bogus" pipeline consisted of warning respondents beforehand that breath samples would be used to determine whether they had smoked. The gain in accuracy came solely from smokers, who were more likely to report that they smoked when they knew their breath would be tested than when they did not know it would be tested. Murray et al. (1987) reviewed 11 studies that used the bogus pipeline procedure to improve adolescents' reports about smoking; 5 of the studies found significant effects for the procedure. Tourangeau, Smith, and Rasinski (1997) examined a range of sensitive topics in a community sample and found significant increases under the bogus pipeline procedure in reports about drinking and illicit drug use. Finally, Wish, Yacoubian, Perez, and Choyka (2000) compared responses from adult arrestees who were asked to provide urine specimens either before or after they answered questions about illicit drug use; there was sharply reduced underreporting of cocaine and marijuana use among those who tested positive for the drugs in the "test first" group (see Yacoubian & Wish, 2001, for a replication). In most of these studies of the bogus (or actual) pipeline, nonresponse is negligible and cannot account for the observed differences between groups. Researchers may be reluctant to use the bogus pipeline procedure when it involves deceiving respondents; we do not know of any national surveys that have attempted to use this method to reduce misreporting.

Forgiving Wording and Other Question Strategies

Surveys often use other tactics in an attempt to improve reporting of sensitive information. Among these are "forgiving" wording of the questions, assurances regarding the confidentiality of the data, and matching of interviewer–respondent demographic characteristics.

Most of the standard texts on writing survey questions recommend "loading" sensitive behavioral questions to encourage respondents to make potentially embarrassing admissions (e.g., Fowler, 1995, pp. 28–45; Sudman & Bradburn, 1982, pp. 71–85).

For example, the question might presuppose the behavior in question ("How many cigarettes do you smoke each day?") or suggest that it is quite common ("Even the calmest parents get angry at their children sometimes. Did your children do anything in the past seven days to make you yourself angry?"). Surprisingly few studies have examined the validity of these recommendations to use "forgiving" wording. Holtgraves, Eck, and Lasky (1997) reported five experiments that varied the wording of questions on sensitive behaviors and found few consistent effects. Their wording manipulations had a much clearer effect on respondents' willingness to admit they did not know much about various attitude issues (such as global warming or the GATT treaty) than on responses to sensitive behavioral questions. Catania et al. (1996) carried out an experiment that produced some evidence for increased reporting (e.g., of extramarital affairs and sexual problems) with forgiving wording of the sensitive questions than with more neutral wording, but Abelson, Loftus, and Greenwald (1992) found no effect for a forgiving preamble (". . . we often find a lot of people were not able to vote because they weren't registered, they were sick, or they just didn't have the time") on responses to a question about voting.

There is some evidence that using *familiar* wording can increase reporting; Bradburn, Sudman, and Associates (1979) found a significant increase in reports about drinking and sexual behaviors from the use of familiar terms in the questions ("having sex") as compared to the more formal standard terms ("sexual intercourse"). In addition, Tourangeau and Smith (1996) found a context effect for reports about sexual behavior. They asked respondents to agree or disagree with a series of statements that expressed either permissive or restrictive views about sexuality ("It is only natural for people who date to become sexual partners" versus "It is wrong for a married person to have sexual relations with someone other than his or her spouse"). Contrary to their hypothesis, Tourangeau and Smith found that respondents reported fewer sexual partners when the questions followed the attitude items expressing permissive views than when they followed the ones expressing restrictive views; however, the mode of data collection had a larger effect on responses to the sex partner questions with the restrictive than with the permissive items, suggesting that the restrictive context items had sensitized respondents to the difference between self- and interviewer administration. Presser (1990) also reported two studies that manipulated the context of a question about voting in an attempt to reduce overreporting; in both cases, reports about voting were unaffected by the prior items.

Two other question-wording strategies are worth noting. In collecting income, financial, and other numerical information, researchers sometimes use a technique called *unfolding brackets* to collect partial information from respondents who are unwilling or unable to provide exact amounts. Item nonrespondents (or, in some cases, all respondents) are asked a series of bracketing questions ("Was the amount more or less than $25,000?", "More or less than $100,000?") that allows the researchers to place the respondent in a broad income or asset category. Heeringa, Hill, and Howell

---

[6] In such cases, of course, the method might better be labeled the *true pipeline*. We follow the usual practice here and do not distinguish versions of the technique that actually can detect false reports from those that cannot.

(1993) reported that this strategy reduced the amount of missing financial data by 60% or more, but Juster and Smith (1997) reported that more than half of those who refused to provide an exact figure also refused to answer the bracketing questions. Apparently, some people are willing to provide vague financial information but not detailed figures, but others are unwilling to provide either sort of information. Hurd (1999) noted another drawback to this approach. He argued that the bracketing questions are subject to acquiescence bias, leading to anchoring effects (with the amount mentioned in the initial bracketing question affecting the final answer). Finally, Press and Tanur (2004) have proposed and tested a method—the respondent-generated intervals approach—in which respondents generate both an answer to a question and upper and lower bounds on that answer (values that there is "almost no chance" that the correct answer falls outside of). They used Bayesian methods to generate point estimates and credibility intervals that are based on both the answers and the upper and lower bounds; they demonstrated that these point estimates and credibility intervals are often an improvement over conventional procedures for eliciting sensitive information (about such topics as grade point averages and SAT scores).

Other Tactics

If respondents misreport because they are worried that the interviewer might disapprove of them, they might be more truthful with interviewers whom they think will be sympathetic. A study by Catania et al. (1996) provides some evidence in favor of this hypothesis. Their experiment randomly assigned some respondents to a same-sex interviewer and others to an opposite-sex interviewer; a third experimental group got to choose the sex of their interviewer. Catania et al. concluded that sex matching produces more accurate reports, but the findings varied across items, and there were interactions that qualify many of the findings. In contrast to these findings on live interviewers, Couper, Singer, and Tourangeau (2004) found no effects of the sex of the voice used in an IVR study nor any evidence of interactions between that variable and the sex of the respondent.

As we noted earlier in our discussion of question sensitivity and nonresponse, many surveys include assurances to the respondents that their data will be kept confidential. This seems to boost response rates when the questions are sensitive and also seems to reduce misreporting (Singer et al., 1995).

Cannell, Miller, and Oksenberg (1981) reviewed several studies that examined various methods for improving survey reporting, including two tactics that are relevant here; instructing respondents to provide exact information and asking them to give a signed pledge to try hard to answer the questions increased the accuracy of their answers. In an era of declining response rates, making added demands on the respondents is a less appealing option than it once was, but at least one national survey—the National Medical Expenditure Survey—used a commitment procedure modeled on the one developed by Cannell et al.

Rasinski, Visser, Zagatsky, and Rickett (2005) investigated an alternative method to increase respondent motivation to answer truthfully. They used a procedure that they thought would implicitly prime the motive to be honest. The participants in their study first completed a task that required them to assess the similarity of the meaning of words. For some of the participants, four of the

target words were related to honesty; for the rest, none of the target words were related to honesty. The participants then completed an ostensibly unrelated questionnaire that included sensitive items about drinking and cheating. In line with the hypothesis of Rasinski et al., the participants who got the target words related to honesty were significantly more likely to report various drinking behaviors than were the participants who got target words unrelated to honesty.

One final method for eliciting sensitive information is often mentioned in survey texts: having respondents complete a self-administered questionnaire and placing their completed questionnaire in a sealed ballot box. The one empirical assessment of this method (Krosnick et al., 2002) indicated that the sealed ballot box does not improve reporting.

Summary

Several techniques consistently reduce socially desirable responding: self-administration of the questions, the randomized response technique, collecting the data in private (or at least with no parents present), the bogus pipeline, and priming the motivation to be honest. These methods seem to reflect two underlying principles. They either reduce the respondent's sense of the presence of another person or affect the respondent's motivation to tell the truth or both. Both self-administration and the randomized response technique ensure that the interviewer (if one is present at all) is unaware of the respondent's answers (or of their significance). The presence of third parties, such as parents, who might disapprove of the respondents or punish him or her seems to reduce respondents' willingness to report sensitive information truthfully; the absence of such third parties promotes truthful reporting. Finally, both the bogus pipeline and the priming procedure used by Rasinski et al. (2005) seem to increase the respondents' motivation to report sensitive information. With the bogus pipeline, this motivational effect is probably conscious and deliberate; with the priming procedure, it is probably unconscious and automatic. Confidentiality assurances also have a small impact on willingness to report and accuracy of reporting, presumably by alleviating respondent concerns that the data will end up in the wrong hands.

How Reporting Errors Arise

In an influential model of the survey response process, Tourangeau (1984) argued that there are four major components to the process of answering survey questions. (For greatly elaborated versions of this model, see Sudman, Bradburn, & Schwarz, 1996; Tourangeau et al., 2000.) Ideally, respondents understand the survey questions the way the researcher intended, retrieve the necessary information, integrate the information they retrieve using an appropriate estimation or judgment strategy, and report their answer without distorting it. In addition, respondents should have taken in (or "encoded") the requested information accurately in the first place. Sensitive questions can affect the accuracy of the answers through their effects on any of these components of the response process.

Different theories of socially desirable responding differ in part in which component they point to as the source of the bias. Paulhus's (1984) notion of self-deception is based on the idea that

some respondents are prone to encode their characteristics as positive, leading to a sincere but inflated opinion of themselves. This locates the source of the bias in the encoding component. Holtgraves (2004) suggested that several other components of the response process may be involved instead; for example, he suggested that some respondents may bypass the retrieval and integration steps altogether, giving whatever answer seems most socially desirable without bothering to consult their actual status on the behavior or trait in question. Another possibility, according to Holtgraves, is that respondents do carry out retrieval, but in a biased way that yields more positive than negative information. If most people have a positive self-image (though not necessarily an inflated one) and if memory search is confirmatory, then this might bias responding in a socially desirable direction (cf. Zuckerman, Knee, Hodgins, & Miyake, 1995). A final possibility is that respondents edit their answers before reporting them. This is the view of Tourangeau et al. (2000, chap. 9), who argued that respondents deliberately alter their answers, largely to avoid embarrassing themselves in front of an interviewer.

Motivated Misreporting

Several lines of evidence converge on the conclusion that the main source of error is deliberate misreporting. First, for many sensitive topics, almost all the reporting errors are in one direction—the socially desirable direction. For example, only about 1.3% of voters reported in the American National Election Studies that they did not vote; by contrast, 21.4% of the nonvoters reported that they voted (Belli et al., 2001; see Table 1). Similarly, few adolescents claim to smoke when they have not, but almost half of those who have smoked deny it (Bauman & Dent, 1982; Murray et al., 1987); the results on reporting of illicit drug use follow the same pattern. If forgetting or misunderstanding of the questions were the main issue with these topics, we would expect to see roughly equal rates of error in both directions. Second, procedures that reduce respondents' motivation to misreport, such as self-administration or the randomized response techniques, sharply affect reports on sensitive topics but have few or relatively small effects on nonsensitive topics (Tourangeau et al., 2000, chap. 10). These methods reduce the risk that the respondent will be embarrassed or lose face with the interviewer. Similarly, methods that increase motivation to tell the truth, such as the bogus pipeline (Tourangeau, Smith, & Rasinski, 1997) or the priming technique used by Rasinski et al. (2005) have greater impact on responses to sensitive items than to nonsensitive items. If respondents had sincere (but inflated) views about themselves, it is not clear why these methods would affect their answers. Third, the changes in reporting produced by the bogus pipeline are restricted to respondents with something to hide (Bauman & Dent, 1982; Murray et al., 1987). Asking a nonsmoker whether they smoke is not very sensitive, because they have little reason to fear embarrassment or punishment if they tell the truth. Similarly, the gains from self-administration seem larger the more sensitive the question (Tourangeau & Yan, in press).

So, much of the misreporting about sensitive topics appears to result from a motivated process. In general, the results on self-administration and the privacy of the setting of the interview suggest that two distinct motives may govern respondents' willingness to report sensitive information truthfully. First, respondents may be reluctant to make sensitive disclosures to an interviewer because they are afraid of embarrassing themselves

(Tourangeau et al., 2000, chap. 9) or of losing face (Holtgraves, Eck, & Lasky, 1997). This motive is triggered whenever an interviewer is aware of the significance of their answers (as in a telephone or face-to-face interview with direct questions). Second, they may be reluctant to reveal information about themselves when bystanders and other third parties may learn of it because they are afraid of the consequences; these latter concerns generally center on authority figures, such as parents (Aquilino et al., 2000) or commanding officers (Rosenfeld et al., 1996).

There is evidence that respondents may edit their answers for other reasons as well. For example, they sometimes seem to tailor their answers to avoid offending the interviewer, giving more pro-feminist responses to female interviewers than to male interviewers (Kane & Macaulay, 1993) or reporting more favorable attitudes towards civil rights to Black interviewers than to White ones (Schuman & Converse, 1971; Schuman & Hatchett, 1976).

Retrieval Bias Versus Response Editing

Motivated misreporting could occur in at least two different ways (Holtgraves, 2004). It could arise at a relatively late stage of the survey response process, after an initial answer has been formulated; that is, respondents could deliberately alter or edit their answers just before they report them. Misreporting could also occur earlier in the response process, with respondents either conducting biased retrieval or skipping the retrieval step entirely. If retrieval were completely skipped, respondents would simply respond by choosing a socially desirable answer. Or, if they did carry out retrieval, respondents might selectively retrieve information that makes them look good. (Schaeffer, 2000, goes even further, arguing that sensitive questions could trigger automatic processes affecting all the major components of the response process; see her Table 7.10.) Holtgraves's (2004) main findings—that heightened social desirability concerns produced the longer response times whether or not respondents answered in a socially desirable direction—tend to support the editing account. If respondents omitted the retrieval step when the question was sensitive, they would presumably answer more quickly rather than more slowly; if they engaged in biased retrieval, then response times would depend on the direction of their answers. Instead, Holtgraves's findings seem to suggest that respondents engaged in an editing process prior to reporting their answers, regardless of whether they ultimately altered their answers in a socially desirable direction (see Experiments 2 and 3; for a related finding, see Paulhus, Graf, & van Selst, 1989).

Attitudes, Traits, and Behaviors

The psychological studies of socially desirable responding tend to focus on misreporting about traits (beginning with Crowne & Marlowe, 1964; Edwards, 1957; and Wiggins, 1964, and continuing with the work by Paulhus, 1984, 2002) and attitudes (see, e.g., the recent outpouring of work on implicit attitude measures, such as the Implicit Attitudes Test [IAT], for assessing racism, sexism, and other socially undesirable attitudes; Greenwald & Banaji, 1995; Greenwald, McGhee, & Schwartz, 1998). By contrast, the survey studies on sensitive questions have focussed on reports about behaviors. It is possible that the sort of editing that leads to misreporting about sensitive behaviors in surveys (like drug use or

sexual behaviors) is less relevant to socially desirable responding on trait or attitudinal measures.

A couple of lines of evidence indicate that the opposite is true—that is, similar processes lead to misreporting for all three types of self-reports. First, at least four experiments have included both standard psychological measures of socially desirability responding (such as impression management scores from the BIDR) and sensitive survey items as outcome variables (Couper et al., 2003, 2004; and Tourangeau et al., 2003, Studies 1 and 2). All four found that changes in mode of data collection and other design variables tend to affect both survey reports and social desirability scale scores in the same way; thus, there seems to be considerable overlap between the processes tapped by the classic psychological measures of socially desirable responding and those responsible for misreporting in surveys. Similarly, the various implicit measures of attitude (see, e.g., Devine, 1989; Dovidio & Fazio, 1992; Greenwald & Banaji, 1995; Greenwald et al., 1998) are thought to reveal more undesirable attitudes than traditional (and explicit) attitude measures because the implicit measures are not susceptible to conscious distortion whereas the explicit measures are. Implicit attitude measures, such as the IAT, assess how quickly respondents can carry out some ostensibly nonattitudinal task, such as identifying or classifying a word; performance on the task is facilitated or inhibited by positive or negative attitudes. (For criticisms of the IAT approach, see Karpinski & Hilton, 2001; Olson & Fazio, 2004.) In addition, respondents report more socially undesirable attitudes measures (such as race prejudice) on explicit measures of these attitudes when the questions are administered under the bogus pipeline than under conventional data collection (Roese & Jamieson, 1993) and when they are self-administered than when they are administered by an interviewer (Krysan, 1998). Misreporting of undesirable attitudes seems to result from the same deliberate distortion or editing of the answers that produces misreporting about behaviors.

Controlled or Automatic Process?

The question remains to what extent this editing process is automatic or controlled. We argued earlier that editing is deliberate, suggesting that the process is at least partly under voluntary control, but it could have some of the other characteristics of automatic processes (e.g., happening wholly or partly outside of awareness or producing little interference with other cognitive processes). Holtgraves (2004) provided some evidence on this issue. He administered the BIDR, a scale that yields separate scores for impression management and self-deception. Holtgraves argued (as did Paulhus, 1984) that impression management is largely a controlled process, whereas self-deception is mostly automatic. He found that respondents high in self-deception responded reliably faster to sensitive items, but not to nonsensitive items, than did those low in self-deception, consistent with the view that high self-deception scores are the outcome of an automatic process. However, he did not find evidence that respondents high in impression management took significantly more time to respond than did those low in impression management, even though they were more likely to respond in a socially desirable manner than their low IM counterparts. This evidence seems to point to the operation of a fast, relatively effortless editing process.

The notion of a well-practiced, partly automatic editing process is also consistent with studies on lying in daily life (DePaulo et al., 2003; DePaulo, Kashy, Kirkenol, Wyer, & Epstein, 1996). Lying is common in everyday life and it seems only slightly more burdensome cognitively than telling the truth—people report that they do not spend much time planning lies and that they regard their everyday lies as of little consequence (DePaulo et al., 1996, 2003).

By contrast, applications of subjective expected utility theory to survey responding (e.g., Nathan, Sirken, Willis, & Esposito, 1990) argue for a more controlled editing process, in which individuals carefully weigh the potential gains and losses from answering truthfully. In the survey context, the possible gains from truthful reporting include approval from the interviewer or the promotion of knowledge about some important issue; potential losses include embarrassment during the interview or negative consequences from the disclosure of the information to third parties (Rasinski, Baldwin, Willis, & Jobe, 1994; see also Schaeffer, 2000, Table 7.11, for a more detailed list of the possible losses from truthful reporting). Rasinski et al. (1994) did a series of experiments using vignettes that described hypothetical survey interviews; the vignettes varied the method of data collection (e.g., self- versus interviewer administration) and social setting (other family members are present or not). Participants rated whether the respondents in the scenarios were likely to tell the truth. The results of these experiments suggest that respondents are sensitive to the risk of disclosure in deciding whether to report accurately, but the studies do not give much indication as to how they arrive at these decisions (Rasinski et al., 1994).

Overall, then, it remains somewhat unclear to what extent the editing of responses to sensitive questions is an automatic or a controlled process.

Misreporting Versus Item Nonresponse

When asked a sensitive question (e.g., a question about illicit drug use in the past year), respondents can choose to (a) give a truthful response, (b) misreport by understating the frequency or amount of their illicit drug use, (c) misreport by completely denying any use of illicit drugs, or (d) refuse to answer the question. There has been little work on how respondents select among the latter three courses of action. Beatty and Herrmann (2002) argued that three factors contribute to item nonresponse in surveys—the cognitive state of the respondents (that is, how much they know), their judgment of the adequacy of what they know (relative to the level of exactness or accuracy the question seems to require), and their communicative intent (that is, their willingness to report). Respondents opt not to answer a question when they do not have the answer, when they have a rough idea but believe that the question asks for an exact response, or when they simply do not want to provide the information. Following Schaeffer (2000, p. 118), we speculate that in many cases survey respondents prefer misreporting to not answering at all, because refusing to answer is often itself revealing—why would one refuse to answer a question about, say, cocaine use if one had not ever used cocaine at all?

Conclusions

According to models of the survey response process (e.g., Tourangeau, Rips, & Rasinski, 2000), response errors in surveys can

occur because respondents misunderstand the questions, cannot retrieve all the relevant information, use inaccurate estimation and judgment strategies, round their answers, or have difficulty mapping them onto one of the response categories. Answers to sensitive questions are subject to these normal sources of reporting errors, but they have an additional problem as well—respondents simply do not want to tell the truth.

This article summarizes the main findings regarding sensitive questions in surveys. There is some evidence that asking sensitive questions lowers response rates and boosts item nonresponse and reporting errors. There is even stronger evidence that misreporting about sensitive topics is quite common in surveys and that the level of misreporting is responsive to features of the survey design. Much of the misreporting on sensitive topics seems not to involve the usual suspects when it comes to reporting error in surveys, but rather results from a more or less deliberate process in which respondents edit their answers before they report them. Consequently, the errors introduced by editing tend to be in the same direction, biasing the estimates from surveys rather than merely increasing their variance. Respondents are less likely to overreport socially desirable behaviors and to underreport socially undesirable ones when the questions are self-administered, when the randomized response technique or bogus pipeline is used, and when the data are collected in private (or at least away from the respondent's parents). Respondents in surveys seem to lie for pretty much the same reasons they lie in everyday life—to avoid embarrassment or possible repercussions from disclosing sensitive information—and procedures that take these motives into account are more likely to elicit accurate answers. The methodological findings suggest that socially desirable responding in surveys is largely contextual, depending both on the facts of the respondent's situation and on features of the data collection situation such as the degree of privacy it offers.

Future survey procedures to encourage honest reporting are likely to involve new forms of computerized self-administration. There is some evidence (though nonsignificant) that computerization increases the reporting of sensitive information in surveys relative to paper questionnaires, though the difference between the two may depend on other variables (such as whether the computer is a stand-alone machine or part of a network). It remains to be seen whether Web administration of the questions will retain the advantages of other forms of computer administration. Moreover, as ever more elaborate interfaces are adopted, computerization may backfire by conveying a sense of social presence. Future research is needed to determine what features of computerized self-administration are likely to encourage or discourage candid reporting.

Even when the questions are self-administered, whether by computer or on paper, many respondents still misreport when they answer sensitive questions. Thus, another topic for future research is the development of additional methods (or new combinations of methods) for increasing truthful reporting. In a period of declining response rates (e.g., de Leeuw & de Heer, 2002; Groves & Couper, 1998), it is likely that respondents will be more reluctant than they used to be to take part in surveys of sensitive topics and that when they do take part, they will be less inclined to reveal embarrassing information about themselves. The need for methods of data collection that elicit accurate information is more urgent than ever.

References

References marked with an asterisk indicate studies included in the meta-analysis on computerized versus paper self-administration. Those marked with a tilde indicate studies included in the meta-analysis on bystander effects. Finally, those marked with a dagger indicate studies included in the meta-analysis on the item count technique.

Abelson, R. P., Loftus, E. F., & Greenwald, A. G. (1992). Attempts to improve the accuracy of self-reports of voting. In J. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys.* (pp. 138–153). New York: Russell Sage Foundation.

†Ahart, A. M., & Sackett, P. R. (2004). A new method of examining relationships between individual difference measures and sensitive behavior criteria: Evaluating the unmatched count technique. *Organizational Research Methods, 7,* 101–114.

~Aquilino, W. S. (1993). Effects of spouse presence during the interview on survey responses concerning marriage. *Public Opinion Quarterly, 57,* 358–376.

Aquilino, W. S. (1994). Interview mode effects in surveys of drug and alcohol use. *Public Opinion Quarterly, 58,* 210–240.

~Aquilino, W. S. (1997). Privacy effects on self-reported drug use: Interactions with survey mode and respondent characteristics. In L. Harrison & A. Hughes (Eds.), *National Institute on Drug Abuse research monograph series: The validity of self-reported drug use. Improving the accuracy of survey estimates* (No. 167, pp. 383–415). Washington, DC: U.S. Department of Health and Human Services, National Institutes of Health.

Aquilino, W. S., & LoScuito, L. A. (1990). Effect of interview mode on self-reported drug use. *Public Opinion Quarterly, 54,* 362–395.

~Aquilino, W. S., Wright, D. L., & Supple, A. J. (2000). Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use and Misuse, 35,* 845–867.

Bassili, J. N. (1996). The how and the why of response latency measurement in telephone surveys. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 319–346). San Francisco: Jossey-Bass.

Bauman, K., & Dent, C. (1982). Influence of an objective measure on self-reports of behavior. *Journal of Applied Psychology, 67,* 623–628.

Beatty, P., & Herrmann, D. (2002). To answer or not to answer: Decision processes related to survey item nonresponse. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 71–86). New York: Wiley.

*Beebe, T. J., Harrison, P. A., McRae, J. A., Anderson, R. E., & Fulkerson, J. A. (1998). An evaluation of computer-assisted self-interviews in a school setting. *Public Opinion Quarterly, 62,* 623–632.

Belli, R. F., Traugott, M. W., & Beckmann, M. N. (2001). What leads to voting overreports? Contrasts of overreporters to validated votes and admitted nonvoters in the American National Election Studies. *Journal of Official Statistics, 17*(4), 479–498.

Berman, J., McCombs, H., & Boruch, R. F. (1977). Notes on the contamination method: Two small experiments in assuring confidentiality of response. *Sociological Methods and Research, 6,* 45–63.

Biemer, P., & Brown, G. (2005). Model-based estimation of drug use prevalence using item count data. *Journal of Official Statistics, 21,* 287–308.

*Booth-Kewley, S., Edwards, J. E., & Rosenfeld, P. (1992). Impression management, social desirability, and computer administration of attitude questionnaires: Does the computer make a difference? *Journal of Applied Psychology, 77,* 562–566.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis* (Version 2.1) [Computer software]. Englewood, NJ: BioStat.

Bradburn, N., Sudman, S., & Associates. (1979). *Improving interview method and questionnaire design.* San Francisco: Jossey-Bass.

Brener, N. D., Eaton, D. K., Kann, L., Grunbaum, J. A., Gross, L. A., Kyle, T. M., et al. (2006). The association of survey setting and mode with self-reported health risk behaviors among high school students. *Public Opinion Quarterly, 70,* 354–374.

Brittingham, A., Tourangeau, R., & Kay, W. (1998). Reports of smoking in a national survey: Self and proxy reports in self- and interviewer-administered questionnaires. *Annals of Epidemiology, 8,* 393–401.

Cannell, C., Miller, P., & Oksenberg, L. (1981). Research on interviewing techniques. In S. Leinhardt (Ed.), *Sociological methodology 1981* (pp. 389–437). San Francisco: Jossey-Bass.

Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., Hauck, W., & Coates, T. J. (1996). Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. *Public Opinion Quarterly, 60,* 345–375.

Catania, J. A., Gibson, D. R., Chitwood, D. D., & Coates, T. J. (1990). Methodological problems in AIDS behavioral research: Influences of measurement error and participation bias in studies of sexual behavior. *Psychological Bulletin, 108,* 339–362.

*Chromy, J., Davis, T., Packer, L., & Gfroerer, J. (2002). Mode effects on substance use measures: Comparison of 1999 CAI and PAPI data. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (pp. 135–159). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Clark, J. P., & Tifft, L. L. (1966). Polygraph and interview validation of self-reported deviant behavior. *American Sociological Review, 31,* 516–523.

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web- or internet-based surveys. *Educational and Psychological Measurement, 60,* 821–836.

Corkrey, R., & Parkinson, L. (2002). A comparison of four computer-based telephone interviewing methods: Getting answers to sensitive questions. *Behavior Research Methods, Instruments, & Computers, 34*(3), 354–363.

Couper, M. P., Singer, E., & Kulka, R. A. (1998). Participation in the 1990 decennial census: Politics, privacy, pressures. *American Politics Quarterly, 26,* 59–80.

Couper, M. P., Singer, E., & Tourangeau, R. (2003). Understanding the effects of audio-CASI on self-reports of sensitive behavior. *Public Opinion Quarterly, 67,* 385–395.

Couper, M. P., Singer, E., & Tourangeau, R. (2004). Does voice matter? An interactive voice response (IVR) experiment. *Journal of Official Statistics, 20,* 551–570.

Crowne, D., & Marlowe, D. (1964). *The approval motive.* New York: John Wiley.

Currivan, D. B., Nyman, A. L., Turner, C. F., & Biener, L. (2004). Does telephone audio computer-assisted self-interviewing improve the accuracy of prevalence estimates of youth smoking: Evidence from the UMASS Tobacco Study? *Public Opinion Quarterly, 68,* 542–564.

Dalton, D. R., Daily, C. M., & Wimbush, J. C. (1997). Collecting "sensitive" data in business ethics research: A case for the unmatched block technique. *Journal of Business Ethics, 16,* 1049–1057.

†Dalton, D. R., Wimbush, J. C., & Daily, C. M. (1994). Using the unmatched count technique (UCT) to estimate base-rates for sensitive behavior. *Personnel Psychology, 47,* 817–828.

de Leeuw, E., & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. Groves, D. Dillman, J. Eltinge, & R. Little (Eds.), *Survey nonresponse* (pp. 41–54). New York: John Wiley.

de Leeuw, E., & van der Zouwen, J. (1988). Data quality in telephone and face to face surveys: A comparative meta-analysis. In R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, & J. Waksberg (Eds.) *Telephone survey methodology* (pp. 283–299). New York: Wiley.

DeMaio, T. J., (1984). Social desirability and survey measurement: A review. In C. F. Turner & E. Martin (Eds.), *Surveying subjective phenomena* (Vol. 2, pp. 257–281). New York: Russell Sage.

DePaulo, B. M., Kashy, D. A., Kirkenol, S. E., Wyer, M. W., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70,* 979–995.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129,* 74–118.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18.

Dillman, D. A., Sinclair, M. D., & Clark, J. R. (1993). Effects of questionnaire length, respondent friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. *Public Opinion Quarterly, 57,* 289–304.

Dovidio, J. F., & Fazio, R. H. (1992). New technologies for the direct and indirect assessment of attitudes. In J. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 204–237). New York: Russell Sage Foundation.

†Droitcour, J., Caspar, R. A., Hubbard, M. L., Parsely, T. L., Visscher, W., & Ezzati, T. M. (1991). The item count technique as a method of indirect questioning: A review of its development and a case study application. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 185–210). New York: Wiley.

Droitcour, J., Larson, E. M., & Scheuren, F. J. (2001). The three card method: Estimation sensitive survey items—with permanent anonymity of response. In *Proceedings of the Section on Survey Research Methods, American Statistical Association.* Alexandria, VA: American Statistical Association.

Duffy, J. C., & Waterton, J. J. (1984). Under-reporting of alcohol consumption in sample surveys: The effect of computer interviewing in fieldwork. *British Journal of Addiction, 79,* 303–308.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455–463.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden.

Epstein, J. F., Barker, P. R., & Kroutil, L. A. (2001). Mode effects in self-reported mental health data. *Public Opinion Quarterly, 65,* 529–549.

Erdman, H., Klein, M. H., & Greist, J. H. (1983). The reliability of a computer interview for drug use/abuse information. *Behavior Research Methods & Instrumentation, 15*(1), 66–68.

*Evan, W. M., & Miller, J. R. (1969). Differential effects on response bias of computer vs. conventional administration of a social science questionnaire: An exploratory methodological experiment. *Behavioral Science, 14,* 216–227.

Fendrich, M., & Johnson, T. P. (2001). Examining prevalence differences in three national surveys of youth: Impact of consent procedures, mode, and editing rules. *Journal of Drug Issues, 31*(3), 615–642.

Fendrich, M., & Vaughn, C. M. (1994). Diminished lifetime substance use over time: An inquiry into differential underreporting. *Public Opinion Quarterly, 58,* 96–123.

Fogg, B. J., & Nass, C. (1997). Silicon sycophants: The effects of computers that flatter. *International Journal of Human-Computer Studies, 46,* 551–561.

Fowler, F. J. (1995). *Improving survey questions: Design and evaluation.* Thousand Oaks, CA: Sage Publications.

Fowler, F. J., & Stringfellow, V. L. (2001). Learning from experience: Estimating teen use of alcohol, cigarettes, and marijuana from three survey protocols. *Journal of Drug Issues, 31,* 643–664.

Fu, H., Darroch, J. E., Henshaw, S. K., & Kolb, E. (1998). Measuring the

extent of abortion underreporting in the 1995 National Survey of Family Growth. *Family Planning Perspectives, 30,* 128–133 & 138.

Fujii, E. T., Hennessy, M., & Mak, J. (1985). An evaluation of the validity and reliability of survey response data on household electricity conservation. *Evaluation Review, 9,* 93–104.

~Gfroerer, J. (1985). Underreporting of drug use by youths resulting form lack of privacy in household interviews. In B. Rouse, N. Kozel, & L. Richards (Eds.), *Self-report methods for estimating drug use: Meeting current challenges to validity* (pp. 22–30). Washington, DC: National Institute on Drug Abuse.

Gfroerer, J., & Hughes, A. (1992). Collecting data on illicit drug use by phone. In C. Turner, J. Lessler, & J. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (DHHS Publication No. ADM 92–1929, pp. 277–295). Washington, DC: U.S. Government Printing Office.

Greenberg, B. G., Abul-Ela, A.-L. A., Simmons, W. R., & Horwitz, D. G. (1969). The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association, 64,* 520–539.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4–27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74,* 1464–1480.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household surveys.* New York: John Wiley.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology.* New York: Wiley.

Groves, R. M., & Kahn, R. (1979). *Surveys by telephone: A national comparison with personal interviews.* New York: Academic Press.

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68,* 2–31.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-salience theory of survey participation: Description and an illustration. *Public Opinion Quarterly, 64,* 125–148.

Guarino, J. A., Hill, J. M., & Woltman, H. F. (2001). *Analysis of the Social Security Number notification component of the Social Security Number, Privacy Attitudes, and Notification experiment.* Washington, DC: U.S. Census Bureau, Planning, Research, and Evaluation Division.

Harrison, L. D. (2001). Understanding the differences in youth drug prevalence rates produced by the MTF, NHSDA, and YRBS studies. *Journal of Drug Issues, 31,* 665–694.

Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review, 43,* 447–462.

Heeringa, S., Hill, D. H., & Howell, D. A. (1993). *Unfolding brackets for reducing item nonresponse in economic surveys* (AHEAD/HRS Report No. 94–029). Ann Arbor, MI: Institute for Social Research.

Henson, R., Cannell, C. F., & Roth, A. (1978). Effects of interview mode on reporting of moods, symptoms, and need for social approval. *Journal of Social Psychology, 105,* 123–129.

Hochstim, J. (1967). A critical comparison of three strategies of collecting data from households. *Journal of the American Statistical Association, 62,* 976–989.

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly, 67,* 79–125.

Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin, 30,* 161–172.

Holtgraves, T., Eck, J., & Lasky, B. (1997). Face management, question

wording, and social desirability. *Journal of Applied Social Psychology, 27,* 1650–1671.

Hughes, A., Chromy, J., Giacoletti, K., & Odom, D. (2002). Impact of interviewer experience on respondent reports of substance use. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (pp. 161–184). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Hurd, M. (1999). Anchoring and acquiescence bias in measuring assets in household surveys. *Journal of Risk and Uncertainty, 19,* 111–136.

Jackson, D. N., & Messick, S. (1961). Acquiescence and desirability as response determinants on the MMPI. *Educational and Psychological Measurement, 21,* 771–790.

Johnson, L. D., & O'Malley, P. M. (1997). The recanting of earlier reported drug use by young adults. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (pp. 59–80). Rockville, MD: National Institute on Drug Abuse.

Johnson, T., Hougland, J., & Clayton, R. (1989). Obtaining reports of sensitive behaviors: A comparison of substance use reports from telephone and face-to-face interviews. *Social Science Quarterly, 70,* 174–183.

Johnson, T., & van de Vijver, F. J. (2002). Social desirability in cross-cultural research. In J. Harness, F. J. van de Vijver, & Mohler, P. (Eds.), *Cross-cultural survey methods* (pp. 193–202). New York: Wiley.

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76,* 349–364.

Jones, E. F., & Forrest, J. D. (1992). Underreporting of abortion in surveys of U.S. women: 1976 to 1988. *Demography, 29,* 113–126.

Junn, J. (2001, April). *The influence of negative political rhetoric: An experimental manipulation of Census 2000 participation.* Paper presented at the 59th Annual National Meeting of the Midwest Political Science Association, Chicago.

Juster, T., & Smith, J. P. (1997). Improving the quality of economic data: Lessons from the HRS and AHEAD. *Journal of the American Statistical Association, 92,* 1268–1278.

Kane, E. W., & Macaulay, L. J. (1993). Interviewer gender and gender attitudes. *Public Opinion Quarterly, 57,* 1–28.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Associations Test. *Journal of Personality and Social Psychology, 81,* 774–788.

*Kiesler, S., & Sproull, L. S. (1986). Response effects in the electronic survey. *Public Opinion Quarterly, 50,* 402–413.

*King, W. C., & Miles, E. W. (1995). A quasi-experimental assessment of the effect of computerizing noncognitive paper-and-pencil measurements: A test of measurement equivalence. *Journal of Applied Psychology, 80,* 643–651.

*Knapp, H., & Kirk, S. A. (2003). Using pencil and paper, Internet and touch-tone phones for self-administered surveys: Does methodology matter? *Computers in Human Behavior, 19,* 117–134.

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W. M., Kopp, R. J., et al. (2002). The impact of "no opinion" response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly, 66,* 371–403.

Krysan, M. (1998). Privacy and the expression of white racial attitudes: A comparison across three contexts. *Public Opinion Quarterly, 62,* 506–544.

†LaBrie, J. W., & Earleywine, M. (2000). Sexual risk behaviors and alcohol: Higher base rates revealed using the unmatched-count technique. *Journal of Sex Research, 37,* 321–326.

*Lautenschlager, G. J., & Flaherty, V. L. (1990). Computer administration of questions: More desirable or more social desirability? *Journal of Applied Psychology, 75,* 310–314.

Lee, R. M. (1993). *Doing research on sensitive topics.* London: Sage Publications.

Lemmens, P., Tan, E. S., & Knibbe, R. A. (1992). Measuring quantity and

frequency of drinking in a general population survey: A comparison of five indices. *Journal of Studies on Alcohol, 53,* 476–486.

Lensvelt-Mulders, G. J. L. M., Hox, J., van der Heijden, P. G. M., & Maas, C. J. M. (2005). Meta-analysis of randomized response research. *Sociological Methods and Research, 33,* 319–348.

*Lessler, J. T., Caspar, R. A., Penne, M. A., & Barker, P. R. (2000). Developing computer assisted interviewing (CAI) for the National Household Survey on Drug Abuse. *Journal of Drug Issues, 30,* 9–34.

Lessler, J. T., & O'Reilly, J. M. (1997). Mode of interview and reporting of sensitive issues: Design and implementation of audio computer-assisted self-interviewing. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (pp. 366–382). Rockville, MD: National Institute on Drug Abuse.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage Publications.

Locander, W. B., Sudman, S., & Bradburn, N. M. (1976). An investigation of interview method, threat, and response distortion. *Journal of the American Statistical Association, 71,* 269–275.

Lombard, M., & Ditton, T. B. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication, 3*(2). Retrieved from http://jcmc.indiana.edu/vol3/issue2/lombard.html

Mangione, T. W., Hingson, R. W., & Barrett, J. (1982). Collecting sensitive data: A comparison of three survey strategies. *Sociological Methods and Research, 10,* 337–346.

*Martin, C. L., & Nagao, D. H. (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology, 74,* 72–80.

Messick, S. (1991). Psychology and the methodology of response styles. In R. E. Snow & D. E. Wiley (Eds.), *Improving inquiry in social science: A volume in honor of Lee J. Cronbach* (pp. 200–221). Hillsdale, NJ: Erlbaum.

Moon, Y. (1998). Impression management in computer-based interviews: The effects of input modality, output modality, and distance. *Public Opinion Quarterly, 62,* 610–622.

Moore, J. C., Stinson, L. L., & Welniak, Jr. E. J. (1999). Income reporting in surveys: Cognitive issues and measurement error. In In M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, & R. Tourangeau (Eds.), *Cognition and survey research* (pp. 155–173). New York: Wiley.

Mosher, W. D., & Duffer, A. P., Jr. (1994, May). *Experiments in survey data collection: The National Survey of Family Growth pretest.* Paper presented at the meeting of the Population Association of America, Miami, FL.

Moskowitz, J. 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,* 565–587.

Mott, F. (1985). *Evaluation of fertility data and preliminary analytic results from the 1983 survey of the National Longitudinal Surveys of Work Experience of Youth.* Columbus: Ohio State University, Center for Human Resources Research.

Murray, D., O'Connell, C., Schmid, L., & Perry, C. (1987). The validity of smoking self-reports by adolescents: A reexamination of the bogus pipeline procedure. *Addictive Behaviors, 12,* 7–15.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies, 45,* 669–678.

Nass, C., Moon, Y., & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. *Journal of Applied Social Psychology, 29,* 1093–1110.

Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology, 27,* 864–876.

Nathan, G., Sirken, M., Willis, G., & Esposito, J. (1990, November).

Laboratory experiments on the cognitive aspects of sensitive questions. Paper presented at the International Conference on Measurement Error in Surveys, Tucson, AZ.

Newman, J. C., DesJarlais, D. C., Turner, C. F., Gribble, J. N., Cooley, P. C., & Paone, D. (2002). The differential effects of face-to-face and computer interview modes. *American Journal of Public Health, 92,* 294–297.

*O'Reilly, J., Hubbard, M., Lessler, J., Biemer, P., & Turner, C. (1994). Audio and video computer assisted self-interviewing: Preliminary tests of new technology for data collection. *Journal of Official Statistics, 10,* 197–214.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Associations Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86,* 653–667.

Parry, H., & Crossley, H. (1950). Validity of responses to survey questions. *Public Opinion Quarterly, 14,* 61–80.

Patrick, D. L., Cheadle, A., Thompson, D. C., Diehr, P., Koepsell, T., & Kinne, S. (1994). The validity of self-reported smoking: A review and meta-analysis. *American Journal of Public Health, 84,* 1086–1093.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598–609.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah, NJ: Erlbaum.

Paulhus, D. L., Graf, P., & van Selst, M. (1989). Attentional load increases the positivity of self-presentation. *Social Cognition, 7,* 389–400.

Philips, D. L., & Clancy, K. J. (1970). Response biases in field studies of mental illness. *American Sociological Review, 36,* 512–514.

Philips, D. L., & Clancy, K. J. (1972). Some effects of "social desirability" in survey studies. *American Journal of Sociology, 77,* 921–940.

Press, S. J., & Tanur, J. M. (2004). An overview of the respondent-generated intervals (RGI) approach to sample surveys. *Journal of Modern Applied Statistical Methods, 3,* 288–304.

Presser, S. (1990). Can changes in context reduce vote overreporting in surveys? *Public Opinion Quarterly, 54,* 586–593.

Presser, S., & Stinson, L. (1998). Data collection mode and social desirability bias in self-reported religious attendance. *American Sociological Review, 63,* 137–145.

Rasinski, K. A., Baldwin, A. K., Willis, G. B., & Jobe, J. B. (1994, August). *Risk and loss perceptions associated with survey reporting of sensitive behaviors.* Paper presented at the annual meeting of the American Statistical Association, Toronto, Canada.

Rasinski, K. A., Visser, P. S., Zagatsky, M., & Rickett, E. M. (2005). Using implicit goal priming to improve the quality of self-report data. *Journal of Experimental Social Psychology, 41,* 321–327.

†Rayburn, N. R., Earleywine, M., & Davison, G. C. (2003a). Base rates of hate crime victimization among college students. *Journal of Interpersonal Violence, 18,* 1209–1211.

†Rayburn, N. R., Earleywine, M., & Davison, G. C. (2003b). An investigation of base rates of anti-gay hate crimes using the unmatched-count technique. *Journal of Aggression, Maltreatment & Trauma, 6,* 137–152.

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interview. *Journal of Applied Psychology, 84,* 754–775.

Richter, L., & Johnson, P. B. (2001). Current methods of assessing substance use: A review of strengths, problems, and developments. *Journal of Drug Issues, 31,* 809–832.

Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin, 114,* 363–375.

Rosenfeld, P., Booth-Kewley, S., Edwards, J., & Thomas, M. (1996). Responses on computer surveys: Impression management, social desir-

ability, and the big brother syndrome. *Computers in Human Behavior, 12,* 263–274.

Sackheim, H. A., & Gur, R. C. (1978). Self-deception, other-deception and consciousness. In G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research* (Vol. 2, pp. 139–197). New York: Plenum Press.

Schaeffer, N. C. (2000). Asking questions about threatening topics: A selective overview. In A. Stone, J. Turkkan, C. Bachrach, V. Cain, J. Jobe, & H. Kurtzman (Eds.), *The science of self-report: Implications for research and practice* (pp. 105–121). Mahwah, NJ: Erlbaum.

Schober, S., Caces, M. F., Pergamit, M., & Branden, L. (1992). Effects of mode of administration on reporting of drug use in the National Longitudinal Survey. In C. Turner, J. Lessler, & J. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (pp. 267–276). Rockville, MD: National Institute on Drug Abuse.

Schuessler, K., Hittle, D., & Cardascia, J. (1978). Measuring responding desirability with attitude-opinion items. *Social Psychology, 41,* 224–235.

Schuman, H., & Converse, J. (1971). The effects of black and white interviewers on black responses in 1968. *Public Opinion Quarterly, 35,* 44–68.

Schuman, H., & Hatchett, S. (1976). White respondents and race-of-interviewer effects. *Public Opinion Quarterly, 39,* 523–528.

~Schutz, C. G., & Chicoat, H. D. (1994). Breach of privacy in surveys on adolescent drug use: A methodological inquiry. *International Journal of Methods in Psychiatric Research, 4,* 183–188.

~Silver, B. D., Abramson, P. R., & Anderson, B. A. (1986). The presence of others and overreporting of voting in American national elections. *Public Opinion Quarterly, 50,* 228–239.

Singer, E., Hippler, H., & Schwarz, N. (1992). Confidentiality assurances in surveys: Reassurance or threat. *International Journal of Public Opinion Research, 4,* 256–268.

Singer, E., Mathiowetz, N., & Couper, M. (1993). The impact of privacy and confidentiality concerns on survey participation: The case of the 1990 U.S. census. *Public Opinion Quarterly, 57,* 465–482.

Singer, E., & Presser, S. (in press). Privacy, confidentiality, and respondent burden as factors in telephone survey nonresponse. In J. Lepkowski, C. Tucker, J. M. Brick, E. de Leeuw, L. Japec, P. J. Lavrakas et al. (Eds.), *Advances in telephone survey methodology.* New York: Wiley.

Singer, E., Van Hoewyk, J., & Neugebauer, R. (2003). Attitudes and behavior: The impact of privacy and confidentiality concerns on participation in the 2000 Census. *Public Opinion Quarterly, 65,* 368–384.

Singer, E., Van Hoewyk, J., & Tourangeau, R. (2001). *Final report on the 1999–2000 surveys of privacy attitudes* (Contract No. 50-YABC-7–66019, Task Order No. 46-YABC-9–00002). Washington, DC: U.S. Census Bureau.

Singer, E., von Thurn, D., & Miller, E. (1995). Confidentiality assurances and response: A quantitative review of the experimental literature. *Public Opinion Quarterly, 59,* 66–77.

Sirken, M. G. (1970). Household surveys with multiplicity. *Journal of the American Statistical Association, 65,* 257–266.

*Skinner, H. A., & Allen, B. A. (1983). Does the computer make a difference? Computerized versus face-to-face versus self-report assessment of alcohol, drug, and tobacco use. *Journal of Consulting and Clinical Psychology, 51,* 267–275.

Smith, T. W. (1992). Discrepancies between men and women in reporting number of sexual partners: A summary from four countries. *Social Biology, 39,* 203–211.

Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K. (1996). When the interface is a face. *Human-Computer Interaction, 11,* 97–124.

Stulginskas, J. V., Verreault, R., & Pless, I., B. (1985). A comparison of observed and reported restraint use by children and adults. *Accident Analysis & Prevention, 17,* 381–386.

Sudman, S. (2001). Examining substance abuse data collection methodologies. *Journal of Drug Issues, 31,* 695–716.

Sudman, S., & Bradburn, N. (1974). *Response effects in surveys: A review and synthesis.* Chicago: Aldine.

Sudman, S., & Bradburn, N. (1982). *Asking questions: A practical guide to questionnaire design.* San Francisco: Jossey-Bass.

Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology.* San Francisco: Jossey-Bass.

Sykes, W., & Collins, M. (1988). Effects of mode of interview: Experiments in the UK. In R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 301–320). New York: Wiley.

~Taietz, P. (1962). Conflicting group norms and the "third" person in the interview. *The American Journal of Sociology, 68,* 97–104.

Tourangeau, R. (1984). Cognitive science and survey methods. In T. Jabine, M. Straf, J. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey design: Building a bridge between disciplines* (pp. 73–100). Washington: National Academy Press.

Tourangeau, R., Couper, M. P., & Steiger, D. M. (2003). Humanizing self-administered surveys: Experiments on social presence in web and IVR surveys. *Computers in Human Behavior, 19,* 1–24.

Tourangeau, R., Rasinski, K., Jobe, J., Smith, T. W., & Pratt, W. (1997). Sources of error in a survey of sexual behavior. *Journal of Official Statistics, 13,* 341–365.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response.* Cambridge, England: Cambridge University Press.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly, 60,* 275–304.

Tourangeau, R., Smith, T. W., & Rasinski, K. (1997). Motivation to report sensitive behaviors in surveys: Evidence from a bogus pipeline experiment. *Journal of Applied Social Psychology, 27,* 209–222.

Tourangeau, R., & Yan, T. (in press). Reporting issues in surveys of drug use. *Substance Use and Misuse.*

Traugott, M. P., & Katosh, J. P. (1979). Response validity in surveys of voting behavior. *Public Opinion Quarterly, 43,* 359–377.

*Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998, May 8). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science, 280,* 867–873.

Turner, C. F., Lessler, J. T., & Devore, J. (1992). Effects of mode of administration and wording on reporting of drug use. In C. Turner, J. Lessler, & J. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (pp. 177–220). Rockville, MD: National Institute on Drug Abuse.

Walker, J., Sproull, L., & Subramani, R. (1994). Using a human face in an interface. *Proceedings of the conference on human factors in computers* (pp. 85–99). Boston: ACM Press.

Warner, S. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association, 60,* 63–69.

Warriner, G. K., McDougall, G. H. G., & Claxton, J. D. (1984). Any data or none at all? Living with inaccuracies in self-reports of residential energy consumption. *Environment and Behavior, 16,* 502–526.

Wiggins, J. S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology, 23,* 419–427.

Wiggins, J. S. (1964). Convergences among stylistic response measures from objective personality tests. *Educational and Psychological Measurement, 24,* 551–562.

†Wimbush, J. C., & Dalton, D. R. (1997). Base rate for employee theft: Convergence of multiple methods. *Journal of Applied Psychology, 82,* 756–763.

Wish, E. D., Hoffman, J. A., & Nemes, S. (1997). The validity of self-reports of drug use at treatment admission and at follow-up: Comparisons with urinalysis and hair assays. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates* (pp. 200–225). Rockville, MD: National Institute on Drug Abuse.

Wish, E. D., Yacoubian, G., Perez, D. M., & Choyka, J. D. (2000, November). *Test first and ask questions afterwards: Improving the validity of self-report drug use by arrestees.* Paper presented at the annual meeting of the American Society of Criminology, San Francisco.

Wolter, K. (1985). *Introduction to variance estimation.* New York: Springer-Verlag.

*Wright, D. L., Aquilino, W. S., & Supple, A. J. (1998). A comparison of computer-assisted and paper-and-pencil self-administrated questionnaires in a survey on smoking, alcohol, and drug use. *Public Opinion Quarterly, 62,* 331–353.

Wyner, G. A. (1980). Response errors in self-reported number of arrests. *Sociological Methods and Research, 9,* 161–177.

Yacoubian, G., & Wish, E. D. (2001, November). *Using instant oral fluid tests to improve the validity of self-report drug use by arrestees.* Paper presented at the annual meeting of the American Society of Criminology, Atlanta, GA.

Zuckerman, M., Knee, C. R., Hodgins, H. S., & Miyake, K. (1995). Hypothesis confirmation: The joint effect of positive test strategy and acquiescent response set. *Journal of Personality and Social Psychology, 68,* 52–60.

New Editors Appointed, 2009–2014

The Publications and Communications Board of the American Psychological Association announces the appointment of six new editors for 6-year terms beginning in 2009. As of January 1, 2008, manuscripts should be directed as follows:

- *Journal of Applied Psychology* (http://www.apa.org/journals/apl), **Steve W. J. Kozlowski, PhD,** Department of Psychology, Michigan State University, East Lansing, MI 48824.
- *Journal of Educational Psychology* (http://www.apa.org/journals/edu), **Arthur C. Graesser, PhD,** Department of Psychology, University of Memphis, 202 Psychology Building, Memphis, TN 38152.
- *Journal of Personality and Social Psychology: Interpersonal Relations and Group Processes* (http://www.apa.org/journals/psp), **Jeffry A. Simpson, PhD,** Department of Psychology, University of Minnesota, 75 East River Road, N394 Elliott Hall, Minneapolis, MN 55455.
- *Psychology of Addictive Behaviors* (http://www.apa.org/journals/adb), **Stephen A. Maisto, PhD,** Department of Psychology, Syracuse University, Syracuse, NY 13244.
- *Behavioral Neuroscience* (http://www.apa.org/journals/bne), **Mark S. Blumberg, PhD,** Department of Psychology, University of Iowa, E11 Seashore Hall, Iowa City, IA 52242.
- *Psychological Bulletin* (http://www.apa.org/journals/bul), **Stephen P. Hinshaw, PhD,** Department of Psychology, University of California, Tolman Hall #1650, Berkeley, CA 94720. (Manuscripts will not be directed to Dr. Hinshaw until July 1, 2008, as Harris Cooper will continue as editor until June 30, 2008.)

Electronic manuscript submission: As of January 1, 2008, manuscripts should be submitted electronically via the journal's Manuscript Submission Portal (see the website listed above with each journal title).

Manuscript submission patterns make the precise date of completion of the 2008 volumes uncertain. Current editors, Sheldon Zedeck, PhD, Karen R. Harris, EdD, John F. Dovidio, PhD, Howard J. Shaffer, PhD, and John F. Disterhoft, PhD, will receive and consider manuscripts through December 31, 2007. Harris Cooper, PhD, will continue to receive manuscripts until June 30, 2008. Should 2008 volumes be completed before that date, manuscripts will be redirected to the new editors for consideration in 2009 volumes.