Original Article

The Effect of Self-Report Data Gathering Technique on Prevalence Estimates of Sensitive Behavior

Franklin J. Boster [1], Allison Z. Shaw[2] and James C. Anthony[3]

[1] Department of Communication, Michigan State University, East Lansing, MI 48824, USA

[2] College of Communication Arts & Sciences, Michigan State University, East Lansing, MI 48824, USA

[3] Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI 48824, USA

Corresponding to
Franklin J. Boster
Department of Communication, 404 Wilson Rd., Michigan State University, East Lansing, MI 48824, USA
Email: boster@msu.edu

ABSTRACT

Despite the fact that obtaining accurate self-reports presents challenges for sensitive topics, investigators often employ them to estimate the prevalence of a variety of sensitive behavior. This study examined the effect of four data collection techniques (face-to-face, computer, randomized response technique, and item count technique) on estimates of the prevalence of five sensitive topics (stealing, anal intercourse, lying, marijuana use, and cheating). Each respondent answered all of the five sensitive items, so that item served as a repeated measure. Each respondent responded using the same data collection technique, so that technique served as an independent groups factor. Data were collected at four locations at a large university in the Midwest United States. Four experimenters, two males and two females, solicited respondents. Only those walking alone were solicited, with a random number (2 to 6 inclusive) determining which of the passersby were approached. Consistent with past research the randomized response technique yielded the highest overall prevalence estimates, although for the most sensitive items estimates were within sampling error of those found in the computer condition. Given the advantages in ease and efficiency associated with collecting self-report data with computers, relative to collecting self-report data using the randomized response technique, these results suggest that collecting sensitive information with computers provides considerable value.

S ome of the most interesting and significant questions regarding human behavior involve highly sensitive topics, such as personal sexual practices or illicit substance use. Due to the private and potentially non-normative or illegal nature of sensitive behavior, obtaining accurate prevalence estimates for them provides challenges. These challenges have consequences. To provide one example, the inability to obtain accurate prevalence estimates for sensitive topics has the potential to mislead scholars and practitioners to believe that health campaigns are ineffective when they are effective, or the reverse, and unnecessary when they are necessary, or the reverse.

Generally, estimates of the prevalence of sensitive behavior are

calculated from self-report data, and the self-report data tend to be responses to questions asked face-to-face (FTF) by an interviewer. Citing Grice's (1975) principles that guide conversation; namely, that speakers are expected to be truthful (quality), informative (quantity), relevant (relevance), and clear (manner), and that listeners expect that they will; Schwarz (1999) presents a compelling argument for the utility of viewing the process of obtaining self-reports as a conversation. The goal of soliciting self-report estimates of the prevalence of any act is a brief conversation between an interviewer and a respondent that elicits a quality response, i.e., one that the respondent believes to be true. Grice's other criteria are less applicable when questions are simple, stated clearly, and require a response of either yes or no.

To produce a truthful response requires that respondents be both willing and able to provide the correct answer (Fazio et al., 1995; Jones & Sigall, 1971), and the content of a sensitive item in particular has the potential to impact respondents' willingness to answer accurately. For items that are *non-sensitive*, simple, stated clearly, and that require a response of yes or no, unwillingness to provide an accurate response is negligible, and thus, contributes trivially to error in a prevalence estimate. In contrast, for *sensitive* items, regardless of simplicity, clarity of expression, and response scale, unwillingness provides a formidable obstacle to obtaining an accurate prevalence estimate (Tourangeau et al., 2000; Zimmerman & Langer, 1995).

Consider those for whom "yes" would constitute the accurate answer to any of the following set of items: "Have you used marijuana in the past 30 days?"; "Have you ever cheated on an examination?"; "Have you ever stolen money from a friend or family member?"; "Have you lied to someone you know, face-to-face, in the past 24 hours?"; "Have you ever engaged in anal sexual intercourse?" When asked any of these questions directly (i.e., face-

to-face) by an interviewer, respondents are likely to experience embarrassment, as embarrassment arises in public, impersonal interactions, when one violates (or admits to violating) social conventions (Keltner & Buswell, 1997). It evokes the concern that others will view you negatively if you have violated, or admitted to violating, these conventions, and thus enhances the likelihood of providing an inaccurate response to these items as an attempt to present oneself more positively to the interviewer.

The context in which sensitive items are posed has the potential to affect responses to sensitive items. When an interviewer asks a target to respond to a sensitive item in a face-to-face (FTF) encounter, confidentiality may be promised (although respondents may be skeptical) but anonymity cannot be assured, as the interviewer can associate a face or name with the response. Thus, respondents who have engaged in the focal sensitive act find themselves facing a conflict between providing an accurate response on one hand and avoiding embarrassment, and perhaps additional negative consequences, on the other (Glynn & Huge, 2007; Johnson & Richter, 2004; Tourangeau & Smith, 1996; Tourangeau et al., 2010). A technique for eliciting self-reports that makes the likelihood of response anonymity very high has the potential to increase response accuracy by reducing or eliminating obstacles such as embarrassment that contribute to unwillingness to respond accurately.

Alternatives to direct FTF encounters that promise more anonymity, such as mail surveys, might be employed, but they often produce lower response rates (e.g., 39.1%, Greenlaw & Brown-Welty, 2009; 31.5%, Kaplowitz et al., 2004). Other alternatives, such as telephone interviews, face challenges both of low response rate and inadequate sampling frame (Chang & Krosnick, 2009; de Leeuw et al., 2007; Link et al, 2008; McCarty, 2003). Although the requirements of Internet access and requisite web navigation skills associated with online surveys, or the practical

difficulties associated with bringing computers into the field pose challenges, Greenlaw and Brown-Welty (2009) have noted the advantages of online computer surveys. They suggest that online computer surveys provide respondents with an enhanced sense of anonymity.

Bringing a laptop computer into the field (COM), asking respondents to answer sensitive items out of view of the interviewer, and having them enter their answers into a database without providing any identifying information can be expected to afford a high, or higher, degree of perceived response anonymity. Consequently, it is expected that COM responses produce more accurate prevalence estimates for sensitive items than FTF responses.

In addition to asking questions FTF, interviewers might employ indirect methods. For example, the randomized response technique (RRT) provides a probability-based algorithm that yields easily obtained prevalence estimates (Warner, 1965). The paired-alternatives (Zimmerman & Langer, 1995) or unrelated questions (Greenberg et al., 1977) versions of this technique require that respondents use a randomizing generator (e.g., flip two coins of different denominations, coin A and coin B, without the experimenter observing the outcomes). After flipping both coins, respondents are instructed to answer only one of two questions, both of which solicit yes/no responses. The result of flipping coin A determines which of the two questions is answered. For example, if the coin A flip resulted in a head, the instruction might be to answer if the coin B flip resulted in a head (yes or no). Alternatively, if the coin A flip resulted in a tail, the respondent would be instructed to answer the question concerning the focal sensitive behavior. Notably, prevalence estimates can be calculated without the subject disclosing the question to which they responded (cf. Tourangeau & Yan, 2007, for algorithm). In a meta-analytic review Lensvelt-Mulders et al.

(2005) provide substantial evidence consistent with enhanced accuracy of RRT estimates (see also, the Tourangeau & Yan, 2007 meta-analysis).

The item count technique (ICT) is another indirect option (Droitcour et al., 1991; Miller, 1984). One of the most commonly used forms of the ICT is the paired lists technique (e.g., Kuklinski et al., 1997; Sniderman & Grob, 1996). This technique requires that the investigator construct two lists (a short list and a long list) of items that are administered to independent groups of subjects. All short list items are non-sensitive and yield yes/no responses. They are selected for their homogeneity; a substantial number of respondents are expected to answer them in the same manner. The long list includes the items on the short list plus one additional sensitive item, also presented in a yes/no response format. Respondents are assigned randomly to receive the long or short list, and are instructed to report the number of items to which they respond yes, but not to disclose the specific questions to which they responded yes. A prevalence estimate of the sensitive behavior is obtained by subtracting the mean number of items endorsed on the short list from the mean number of items endorsed on the long list.

Relatively little research has been done to examine the effectiveness of the ICT, and results have been inconsistent. Tourangeau and Yan's (2007) meta-analysis of the comparative effectiveness of the ICT and direct questioning demonstrated that the ICT provided higher estimates of sensitive behavior (e.g., fraud) than direct questioning, although the estimates were within sampling error of each other (see Zimmerman and Langer (1995) for additional conflicting findings). The fact that the ICT-direct questioning difference was not statistically significant arises from the relatively underpowered test, there being only seven studies, and the fact that the results exhibited

substantial heterogeneity as a consequence of one, large sample study in which the effect was both negative and ample (see Table 9, p. 873).

Both of these indirect methods have an important limitation relative to a FTF self-report or an anonymous response entered into a computer data base. The indirect methods are relatively inefficient means of collecting information regarding the prevalence of sensitive action. Specifically, to obtain a sample of size N for the RRT technique requires 2N respondents because the initial coin flip is likely to yield answers to the sensitive item for only one-half of the 2N sample. The ICT technique has the same requirement because only N of the 2N respondents will respond to the long list. Moreover, because responses to a specific sensitive item cannot be linked with other individual difference measures (but see de Jong et al., 2012), certain desirable bivariate and multivariate analyses may be impossible to perform.

Given the introduction of the indirect methods, matters of the definition and measurement of sensitivity may be offered. An item may be thought of as more highly sensitive to the extent that it solicits responses for which respondents know the correct answer but are reluctant or resistant to report it. The degree of sensitivity of an item could be measured in at least two ways. First, ceteris paribus, the lower the prevalence estimate of the item, the higher the likelihood that performing the focal act is non-normative and, hence, the higher is its sensitivity. If S denotes sensitivity, and p_x denotes the proportion of persons endorsing the item (calculated across all methods of data collection), then $S_{1x} = 1 - p_x$ denotes the proportion of persons not endorsing the item. Thus, the higher S_1, ceteris paribus, the more sensitive is the item.

Of course, if a sample of people were asked if they had won a Nobel Prize, the resulting prevalence would be meager and S_1 would approximate 1.0 closely. Thus, not all low prevalence items are sensitive; low prevalence may also indicate scarcity.[1] The RRT is expected to produce the highest prevalence estimate for any item for which sensitivity is expressed by underreporting (see Lensvelt-Mulders et al., 2005; Tourangeau & Yan, 2007). Although the RRT may not provide a perfectly accurate prevalence estimate, it is expected to produce the least underreported estimate. And because FTF is expected to produce the lowest prevalence estimate for any given item, a second indicator of sensitivity is the difference in the prevalence estimate between those reporting via the RRT and FTF corrected for how large a difference is possible, i.e., $S_{2x} = (p_{RRT} - p_{FTF})/p_{RRT}$, where p_{RRT} denotes the prevalence estimate of the RRT, p_{FTF} denotes the prevalence estimate obtain from face-to-face solicitation, and S denotes sensitivity.

In the experiment reported subsequently, respondents were asked to answer a series of relatively sensitive items, yet items that varied somewhat in the magnitude of their sensitivity. These self-reports were solicited by FTF, COM, ICT, or RRT. *It was expected that participants would provide lower prevalence estimates in the FTF condition than in conditions that provide greater anonymity, COM and RRT. Consistent with past data, it was also expected that the RRT would generate the highest prevalence estimates. Finally, it was expected that because of comparable perceived anonymity in the two conditions that prevalence estimates collected by COM would be similar to the RRT estimates.* Given the inconsistency of past results, no conjecture concerning ICT is advanced. Instead, this condition was included to provide additional data

---

[1] The sensitivity of some items may be manifested by over reporting, e.g., always washing one's hands after micturition or asking male participants if they did or did not have sexual intercourse prior to their 16th birthday. Such items can be reflected, i.e., reverse coded. The indicators employed in the experiment described in this manuscript all involve sensitivity that is manifested by under reporting.

pertinent to the effectiveness of this infrequently employed technique.

METHOD

Subjects

Data were collected from a convenience sample of 720 students enrolled at a large Midwestern university. Ss were recruited by one of four Es at one of four campus locations. The sample was 48.3% male, which approximates closely the proportion of males attending this university (approximately 47%). The mean age was 21.16 years (SD = 3.09), which approximates closely the mean age of undergraduate students at this university (M = 20.0). A large majority of Ss were Caucasian (78.2%), which is characteristic of undergraduate students at this university (77%), with minority representation approximating closely that of the university population.[2]

Design

An independent groups design was used to examine differences among FTF, ICT, RRT, and COM methods of collecting sensitive information. Respondents were assigned randomly to a condition of the independent groups factor, each responding to each of the five sensitive items. The FTF and COM conditions required 120 Ss to obtain 120 data points; the ICT and RRT conditions required 240 Ss to obtained the requisite 120 data points. In the FTF and COM conditions any given S's responses can be compared across items, but in the ICT and RRT conditions any given S's response to an item is unknown, and thus, cannot be compared across items. Consequently, item cannot be

treated as a repeated measure, so that data analyses are restricted to comparing responses to each item separately across the four data collection techniques.

FTF condition respondents were solicited orally to participate in an experiment concerning the sensitivity of questions. Respondents were promised confidentiality, reminded that their participation was voluntary, and asked not to provide identifying information. As in all four conditions those who agreed to participate were asked some demographic questions followed by a series of yes/no items that included those both relatively high and relatively low in sensitivity. Upon completion, respondents in all conditions were thanked for their participation. In all conditions those who did not agree to participate were thanked for listening to the request.[3] The order in which the sensitive items were asked was assigned randomly prior to data collection, all respondents receiving the same order of item presentation determined by the initial random assignment.

ICT condition respondents were solicited in the same manner as the FTF condition. Those who agreed to participate were handed either a stack of five short lists (5 questions per card) or five long lists (6 questions per card), asked to examine each item list, and asked to report the number of items to which they answered "yes." The assignment of respondents to the short list or long list was made randomly prior to the start of data collection with the constraint that there were an equal number of participants responding to each list.

RRT condition respondents were solicited in the same manner as in the FTF condition. Those who agreed to participate were given two coins and a set of six cards that instructed them to flip each coin and answer the question indicated. To ensure that respondents understood the procedure,

---

[2] Respondents could and did check multiple categories. The proportion identifying as Caucasian is calculated by summing all respondents who checked that category and dividing by 720.

[3] Although some respondents declined participation, none of them declined after receiving instructions. One female respondent in the FTF condition declined to answer the anal intercourse item and her data were discarded.

the first card was always a practice card. When the experimenter believed that the respondent understood the procedure, the respondent was instructed to move on to the second card.

COM condition respondents were solicited in the same manner as in the FTF condition. They were asked to answer the questions on a laptop computer provided by the experimenter. It was clear to the respondents that this procedure insured that their responses could not be matched with any type of identifying information; the IP address being laptop specific, it was the same for all respondents and could not be employed to identify any individual.

Instrumentation

Five sensitive items were employed. They were phrased as follows: "Have you ever stolen money from a friend or family member?"; "Have you used marijuana in the past 30 days?"; "Have you ever cheated on an examination?"; "Have you ever engaged in anal sexual intercourse?"; "Have you lied to someone you know, face-to-face, in the past 24 hours?" This order of presentation was used for all respondents in all conditions of the experiment.

Procedure

Ss' participation was obtained by soliciting at each of four diverse campus locations. A total of 180 data points was collected at each of the four locations. Four experimenters (two males) collected the data. The orders of the experimental conditions and location of data collection were assigned randomly prior to data collection. A random digit (between 2 and 6) was selected to determine which passersby would be approached and asked to participate. Passersby were only approached if they were walking alone.

RESULTS

Table 1 presents the prevalence estimates partitioned by condition (i.e., data collection technique) and sensitivity indices for each item. Analyses performed to examine the effects of experimenters and locations, as well as their interactions with data collection technique, yielded no evidence of substantial effects; thus, these factors are omitted in the results presented subsequently.

Observing the sensitivity indices, especially S_2, indicates that engaging in anal intercourse and stealing from a friend or family member are the most sensitive of these actions, particularly noticeable by their relative values on S_2. Cheating is the least sensitive item, particularly noticeable by its relative value on S_1.[4]

Table 1. *Proportion of Subjects Indicting That They Have in Sensitive Behavior Partitioned by Self-Report Technique,* N = 480

Item	FTF	COM	ICT	RRT	S_1	S_2	MEAN
Anal	.09	.18	.00	.20	.88	.54	.71
Stealing	.12	.20	.13	.27	.82	.56	.69
Lying	.19	.20	.20	.26	.79	.26	.53
Marijuana	.21	.17	.14	.27	.80	.22	.51
Cheating	.38	.53	.48	.48	.53	.21	.37

---

[4] The correlation between S_1 and S_2 was .62. Thus, standardized item alpha for the sum of these two measures is estimated to be .77. Because sensitive items in which underreporting is expected were selected for this experiment, the correlation between the indices is attenuated as a result of restriction in range. Consequently, the estimated reliability is as well. Nevertheless, these data indicate clearly that two of the items (anal intercourse and stealing) are relatively sensitive and one is relatively insensitive (cheating).

Observing the prevalence estimates it is clear that responses to the anal sexual intercourse item varied by technique, $\chi^2 (df = 3, n = 480) = 28.63$, $p < .001$. In large part this effect is driven by the estimated frequency of zero in the ICT condition. Ignoring the ICT condition and focusing only on the other three conditions indicates that the RRT and COM treatments yield substantially higher prevalence estimates than the FTF treatment, the RRT and COM estimates not being substantially different from one another. Combining the COM and RRT conditions and contrasting them with the FTF condition indicates that the former exceed the latter substantially, $\chi^2 (df = 1, n = 360) = 5.59, p = .018, r = .13, OR = 2.29$.[5]

Prevalence estimates for the stealing item also varied substantially by technique, $\chi^2 (df = 3, n = 480) = 11.50, p = .009$. Table 1 indicates that this effect results from the RRT and COM treatments yielding substantially higher prevalence estimates (.23) than the FTF and ICT treatments (.13), $\chi^2 (df = 1, n = 480) = 9.58, p = .002, r = .14, OR = 2.13$, the RRT and COM treatments being within sampling error of one another as were the FTF and ICT treatments.

As Table 1 indicates prevalence estimates varied little across techniques for the lying item, and the omnibus χ^2 test did not allow the null hypothesis to be rejected, $\chi^2 (df = 3, n = 480) = 2.04, p = .564$. Although it may appear that the prevalence estimate in the RRT condition exceeds those obtained by combining the other three conditions, the resulting difference remains within sampling error of zero.

The omnibus χ^2 test of the effect of the technique induction did not allow the null hypothesis to be rejected for the marijuana use item, $\chi^2 (df = 3, n = 480) = 6.83, p = .078$. On the other hand, observing Table 1 suggests that the RRT yielded

higher prevalence estimates compared with the combination of the other three techniques, $\chi^2 (df = 1, n = 480) = 5.10, p = .024, r = .10, OR = 1.75$. Although it is less extreme than the anal intercourse item, this result is driven by the low estimate in the ICT condition. Put differently, the null hypothesis cannot be rejected when contrasting the RRT condition estimate with the FTF estimate, the RRT estimate with the COM estimate, or the RRT estimate with the FTF/COM combined estimate.

For the cheating item the omnibus X^2 test produced no evidence of statistically significant differences among techniques, $\chi^2 (df = 3, n = 480) = 5.65, p = .13$. Examining Table 1, however, suggests that the FTF treatment differs modestly from the combined prevalence estimate of the other three conditions, $\chi^2 (df = 1, n = 480) = 4.69, p = .03, r = .10, OR = 1.59$.

DISCUSSION

The focus of this experiment was comparing the prevalence estimates of sensitive action obtained by different self-report techniques. At least five features of the results are noteworthy.

First, despite the fact that all items appear sensitive, there are substantial differences among them on this dimension. Anal intercourse and stealing from a friend or family member are more sensitive than lying to someone face to face within the past 24 hours or using marijuana within the last 30 days, and lying to someone face to face within the past 24 hours and using marijuana within the last 30 days are, in turn, less sensitive than cheating on an examination.

Second, different prevalence estimates across techniques were observed for several items.

---

5 Given the modest correlation of .13, objection may be raised to the term "substantially." But, given that the prevalence estimate in the combined COM and RRT conditions was .19, the maximum correlation possible (i.e., with zero frequency in the FTF condition) is .27. Thus, the observed coefficient is more than 50% as large as it could be given that the COM/RRT prevalence estimate exceeds the FTF prevalence estimate. The same point applies to other effect size estimates in this data set.

Moreover, these differences were most pronounced for the most sensitive items (anal intercourse and stealing), and, arguably, for the least sensitive of these items (cheating on an examination). This outcome suggests that the effect of item sensitivity and type of data gathering technique may combine non-additively to affect prevalence estimates. But, because of the inability to identifying any one person's response when using the RRT and ICT, a formal test for non-additivity could not be performed.

Third, consistent with the prediction derived from Grice's (1975) conversational implicatures, and from the possibility that embarrassment affects responses to sensitive items, in the main, estimates generated in the FTF condition produced relatively low prevalence estimates. Once again, this pattern was most pronounced for the most sensitive items, and, arguably, for the least sensitive item.

Fourth, there was no evidence that estimates in the COM condition differed markedly from those in the RRT condition. It is fallacious to affirm the null hypothesis and conclude that these two techniques produce no differences in prevalence estimates of sensitive items. Nevertheless, even if the RRT produces slightly higher, and hence presumably more accurate prevalence estimates, given the advantages of sampling efficiency and the ability to link responses with individual participants, using a computer in the field may be a particularly effective, and preferable, data gathering device when collecting sensitive information.

Fifth, the results for the ICT varied considerably relative to the other techniques. The ICT estimates for the most sensitive items were relatively low; whereas, for the remaining three items they were more similar to COM and RRT estimates. The vagaries of the ICT estimates suggest that they may be very much context dependent. Particularly when demographic differences are of interest, the filler items (i.e., the short list) may be endorsed differentially by different demographic groups, giving rise to specious non-additive effects. For example, filler items such as, "have you ever baked a cake" or "have you ever changed a car tire" are likely endorsed with different prevalence by females and males. Including such items on the short list could have a substantial and differential impact on ICT estimates for females and males, and if random assignment should result in a substantially different proportion of males and females for a particular item, mean prevalence estimates could be affected substantially as well. Subsequent research employing the ICT would benefit from pilot studies that examine short lists for focal demographic differences prior to launching experiments that probe differences in the sensitive action of interest.

In addition to examining differences in prevalence estimates as a function of technique and item, it is reasonable to raise the question of the accuracy of these techniques as estimates of the focal population parameter(s). Because these parameters are not known with certainty or even high confidence, simple comparisons are impossible. Nevertheless, the estimates obtained in this experiment can be compared with estimates obtained from other well conducted surveys and experiments. This strategy is applied in subsequent paragraphs as a method of assessing the accuracy of the estimates presented in this manuscript.

No estimate of the prevalence of stealing from a friend or family member could be identified. Nevertheless, employing FTF and telephone interviews Blanco et al. (2008) report that 11.3% of the population said that they had engaged in one form of cheating, shoplifting, at least once in their lifetime. Although the mean prevalence reported in Table 1 for the stealing item exceeds this figure (17.5%), it is notable that the technique most similar to that employed by Blanco et al. (2008), the FTF condition, produced an almost identical figure (.117). The ICT also produced a very similar estimate. In contrast, the COM

and RRT estimates are considerably higher and perhaps inconsistent with Blanco et al. (2008), particularly if people are more reluctant to steal from people they know, and presumably like, than from an impersonal source, such as a business. On the other hand, there are likely more opportunities to steal from friends or family, and the legal sanctions of being caught are considerably less harsh. These social conditions might make the prevalence of such acts exceed those of Blanco et al. (2008). Hence, these higher estimates may well be more accurate than those produced in the FTF and ICT conditions.

The National Youth Risk Behavior Survey is a biannual national survey of adolescents (grades nine to 12) conducted for the purpose of assessing risk behavior. Compared to 19.7% of adolescents in the 2007 National Youth Risk Behavior Survey (CDC, 2007), this experiment found that 20% of subjects reported having used marijuana within the last 30 days. Although the samples vary slightly in mean age, the findings are very similar. As with the stealing item the RRT estimates may be more accurate than those generated by other techniques.

Cheating estimates exhibit substantial variation across studies. For example, Vandehey et al. (2007) report that 20.9% of their respondents indicated that they had cheated on an examination, but Davis et al. (1992) estimate that 64% of their respondents indicated that they cheated on an examination. The reason(s) for such a substantial difference is (are) unknown; it may simply reflect differences in the student bodies of small private liberal arts colleges and large public universities. Notably, the estimate obtained in this experiment (46.7%), indeed the estimates for all of the techniques employed in this experiment, fall between these two extreme figures.

Leichliter et al. (2007) examined the prevalence of heterosexual anal sexual intercourse. Sampling people ranging from 15 to 44 years of age, overall they found that 32% of the sample had engaged in heterosexual anal sexual intercourse. Of those

between the ages of 15 to 19, however, only 10.9% indicated having had heterosexual anal sexual intercourse in their lifetime, and 29.6% of those between the ages of 20 to 24 indicated having engaged in this activity. In the data presented in this manuscript mean age was 21.6 years, so based on the Leichliter et al. data one would expect a prevalence estimate in the range of 10.9% to 29.6%. Moreover, given that approximately 71% of the sample for this experiment was 21 years old or younger, to be consistent with the Leichliter et al. data the estimate would be expected to be somewhat closer to the lower of the two figures. Thus, the overall estimate (13.5%) might replicate the Leichliter et al. result. On the other hand, substantial differences across techniques were observed, and the higher estimates obtained in the COM and RRT conditions are substantially more likely that the low FTF and ICT estimates to replicate Leichliter et al.

Nevertheless, an important difference in the phrasing of the two items renders comparison difficult. Leichliter et al. (2007) specified that the anal sexual act must be heterosexual; in this experiment that specification was absent. Such a difference might well alter estimates in ways making comparisons unsound.

Results of this study suggest that 21.9% of subjects reported having lied to a known other within the past 24 hours. This estimate is substantially lower than the 40.1% that Serota et al. (2009) obtained for the prevalence of lying, in general within the past 24 hours. This finding could indicate a failure to replicate, or it could suggest that a difference in question wording (known other v. other) affected the estimates, i.e., that people lie more frequently to others, both known others and strangers, than only to known others.

In addition to the accuracy of the estimates generated by the techniques employed in this experiment the issue of the generalizability of the results requires subsequent investigation, and to be thorough the generalizability of a number

of factors must be addressed. For example, five items were used in this experiment, and the normative response for each of these items is "no." In addition to sampling sensitive items in addition to these five, it would be profitable to estimate the prevalence of items for which the normative response is to say "yes," so that there is a tendency to over report. Examples might include (1) sexual activity for men (have you had sexual relations with more than X number of partners), weight for women ("do you weigh less than Y pounds"), or items that likely pertain to both sexes ("do you always wash your hands after micturition").

Another important factor is the demographic composition of the respondents (Sears, 1986). Although college-aged students are a population worthy of study in their own right, a richer description of the prevalence of various sensitive actions requires obtaining a sample that is more heterogeneous on a number of factors, as well as requiring changes in the venue of data collection. An important implication of the Leichliter et al. (2007) survey is that some dimensions of individual differences, be they demographic or psychographic, may have a substantial effect on the value of the prevalence estimate obtained. Such a sample would pose a particular challenge for the ICT. Developing short lists that are comparable across a heterogeneous sample is daunting.

Probing the extent to which the results generalize across different instantiations of the techniques, as well as examining the relative prevalence estimates of alternative techniques, is yet another challenge to the generalizability of the results. For example, collecting sensitive data FTF might be done via interview or questionnaire. Collecting sensitive data via computer could be accomplished by handing the respondent a computer, as in this experiment, or by providing a link to an online survey. ICT data could be collected using short lists with varying content, using short and long lists of varying size, and by varying several other features of the technique. Warner (1965) and others have suggested numerous methods for obtaining RRT data. Subsequent experiments would profit by comparing the relative magnitude of the variance in prevalence between techniques with the magnitude of the variance within instantiations of the same technique.

Additionally, the same order of item presentation was employed throughout the experiment so as to lighten the cognitive load required for those collecting the data. Notably however, if order effects exist, then they would introduce error into the prevalence estimates. The assessment of such a possibility awaits future research.

Finally, to facilitate interpreting these data two measures of sensitivity were employed. The correlation between them was substantial ($r = .62$) so that if they are summed as in the last column of Table 1 the resulting reliability estimate (standardized item alpha) is .77. This estimate is likely attenuated because only relatively sensitive items were employed in this experiment, so that the variances in these measures are attenuated and so that the subsequent correlation and reliability estimates are attenuated because of range restriction.

Nevertheless, the focus on estimating relative item sensitivity holds promise for at least two reasons. First, if a highly valid and highly reliable sensitivity index can be constructed, then norms can be constructed to promote future research concerning sensitive topics. Second, the indices developed in this manuscript may generate other indices that either supplement, correct, or extend S_1 and S_2.

REFERENCES

Blanco, C., Grant, J., Petry, N. M., Simpson. H. B., Alegria, A., Liu, S.M., & Hasin, D. (2008). Prevalence and correlates of shoplifting in the United States: Results from the National

Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *American Journal of Psychiatry, 165*(7), 905-913. https://doi.org/10.1176/appi.ajp.2008.07101660

CDC (Center for Disease Control and Prevention). (2007). *The youth risk behavior surveillance system*. U.S. Government Printing Office.

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly, 73*(4), 641-678. https://doi.org/10.1093/poq/nfp075

Davis., S. F., Grover, C. A., Becker, A. H., & McGregor, L. N. (1992). Academic dishonesty: Prevalence, determinants, techniques, and punishments. *Teaching of Psychology, 19*(1), 16-20. https://doi.org/10.1207/s15328023top1901_3

de Jong, M.G., Pieters, R., & Stremersch, S. (2012). Analysis of sensitive questions across cultures: An application of multigroup item random response theory to sexual attitudes and behavior. *Journal of Personality and Social Psychology, 103*(3), 543-564. https://doi.org/10.1037/a0029394

de Leeuw, E., Callegaro, M., Hox, J., Korendijk, E., & Lensvelt-Mulders, G. (2007). The influence of advance letters on response in telephone surveys: A meta-analysis. *Public Opinion Quarterly, 71*(3), 413-443. https://doi.org/10.1093/poq/nfm014

Droitcour, J., Caspar, R. A., Hubbard, M. L., Parsley, 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. P. Biemer (Ed.), *Measurement errors in surveys* (pp. 185-210). John Wiley & Sons.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*(6),

1013-1027. https://doi.org/10.1037/0022-3514.69.6.1013

Glynn, C. J., & Huge, M. E. (2007). Opinions as norms: Applying a return potential model to the study of communication behaviors. *Communication Research, 34*(5), 548-568. https://doi.org/10.1177/0093650207305236

Greenberg, B. G., Kuebler, R. R., Abernathy, J. R., & Horvitz, D. G. (1977). Respondent hazards in the unrelated question randomized response model. *Journal of Statistical Planning and Inference, 1*(1), 53-60. https://doi.org/10.1016/0378-3758(77)90005-2

Greenlaw, C., & Brown-Welty, S. (2009). A comparison of web-based and paper-based survey methods: Testing assumptions of survey mode and response cost. *Evaluation Review, 33*(5), 464-480. https://doi.org/10.1177/0193841x09340214

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41-58). Academic Press.

Johnson, P. B., & Richter, L. (2004). Research note: What if we're wrong: Some possible implications of systematic distortions in adolescents' self-reports of sensitive behaviors. *Journal of Drug Issues, 34*(4), 951-970. https://doi.org/10.1177/002204260403400412

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76*(5), 349-364. https://doi.org/10.1037/h0031617

Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of web and mail survey response rates. *Public Opinion Quarterly, 68*(1), 94-101. https://doi.org/10.1093/poq/nfh006

Keltner, D., & Buswell, B. N. (1997). Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin, 122*(3), 250-270.

https://doi.org/10.1037/0033-2909.122.3.250

Kuklinski, J. H., Sniderman, P. M., Knight, K., Piazza, T., Tetlock, P. E., Lawrence, G. R., & Mellers, B. (1997). Racial prejudice and attitudes toward affirmative action. *American Journal of Political Science, 41*(2), 402-419. https://doi.org/10.2307/2111770

Leichliter, J. S., Chandra, A., Liddon, N., Fenton, K. A., & Aral, S. O. (2007). Prevalence and correlates of heterosexual anal and oral sex in adolescents and adults in the United States. *The Journal of Infectious Diseases, 196*(12), 1852-1859.
https://doi.org/10.1086/522867

Lensvelt-Mulders, G. J. L. M., van der Heijden, P. G. M., & Maas, C. J. M. (2005). Meta-analysis of random response research: Thirty-five years of validation. *Sociological Methods and Research, 33*(3), 319-348.
https://doi.org/10.1177/0049124104268664

Link, M. W., Battaglia, M. P., Frankel, M. R., Osborn, L., & Mokdad, A. H. (2008). A comparison of address-based sampling (ABS) versus random-digit dialing (RDD) for general population surveys. *Public Opinion Quarterly, 72*(1), 6-27.
https://doi.org/10.1093/poq/nfn003

McCarty, C. (2003). Differences in response rate using most recent versus final dispositions in telephone surveys. *Public Opinion Quarterly, 67*(3), 396-406.
https://doi.org/10.1086/377243

Miller, J. D. (1984). *A new survey technique for studying deviant behavior*. [Doctoral dissertation]. The George Washington University.

Schwarz, N. (1999). Self-reports: How do questions shape the answers. *American Psychologist, 54*(2), 93-105.
https://doi.org/10.1037/0003-066x.54.2.93

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology, 51*(3), 515-530.

https://doi.org/10.1037/0022-3514.51.3.515

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in America: Three studies of self-reported lies. *Human Communication Research, 36*(1), 1-25.
https://doi.org/10.1111/j.1468-2958.2009.01366.x

Sniderman, P. M., & Grob, D. B. (1996). Innovations in experimental design in attitude surveys. *Annual Review of Sociology, 22*(1), 377-399.
https://doi.org/10.1146/annurev.soc.22.1.377

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly, 60*(2), 275-304.
https://doi.org/10.1086/297751

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859-883.
https://doi.org/10.1037/0033-2909.133.5.859

Tourangeau, R., Groves, R. M., & Redline, C. D. (2010). Sensitive topics and reluctant respondents: Demonstrating a link between nonresponse bias and measurement error. *Public Opinion Quarterly, 74*(3), 413-432.
https://doi.org/10.1093/poq/nfq004

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

Vandehey, M. A., Diekhoff, G. M., & LaBeff, E. E. (2007). College cheating: A twenty-year follow-up and the addition of an honor code. *Journal of College Student Development, 48*(4), 468-480.
https://doi.org/10.1353/csd.2007.0043

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association, 60*(309), 63-69.
https://doi.org/10.1080/01621459.1965.10480775

Zimmerman, R. S., & Langer, L. M. (1995).

Improving estimates of prevalence rates of sensitive behaviors: The randomized lists technique and consideration of self-reported honesty. *The Journal of Sex Research, 32*(2), 107-117. https://doi.org/10.1080/00224499509551781