

## Effects of Interview Mode on Measuring Depression in Younger Adults

*William S. Aquilino<sup>1</sup>*

Survey mode effects on self-reported depression as measured by the CES-D were estimated in an urban-suburban sample of 18–45 year-old male and female respondents. In this field experiment, respondents selected through an area probability sample were randomly assigned to one of three interview modes: self-administered, face-to-face, or telephone. There were no differences in depression scores between telephone and face-to-face modes. The self-administered mode yielded higher CES-D scores than either of the interviewer-administered modes. Results suggest that the degree of response anonymity during the interview influences respondents' willingness to reveal sensitive, personal information. A concurrent RDD sample of the same population also produced CES-D estimates significantly lower than the self-administered mode but equivalent to the face-to-face interview.

*Key words:* Survey mode; sensitive questions; self-administered questionnaire; social desirability bias.

### 1. Introduction

The Center for Epidemiological Studies-Depression Scale (CES-D; Radloff 1977) is a frequently used measure of psychological distress in surveys concerning health and well-being (Andresen et al. 1994; Liang et al. 1989; Newmann 1986; 1989). Full or partial versions of the instrument have been incorporated in many large-scale surveys such as the U.S. National Survey of Families and Households (NSFH) and innumerable small-scale studies. Although originally designed for use in general population surveys involving face-to-face interviews, the CES-D has been administered in a variety of survey modes, including telephone and self-administered surveys. A critical methodological question is whether alternate interview modes yield comparable data on sensitive measures of health and well-being. This article explores the degree to which responses to the CES-D vary among face-to-face, telephone, and self-administered survey modes.

### 2. Mode Effects in Sensitive Interviews

The 20-item CES-D elicits sensitive information on personal well-being, including feelings of depressed mood, sadness, worthlessness, guilt, hopelessness, and somatic symptoms

<sup>1</sup> Dept. of Child and Family Studies, 1300 Linden Drive, University of Wisconsin, Madison, WI 53706, U.S.A.  
**Acknowledgments:** This research was supported by a grant from the National Institute on Drug Abuse (R01-DA06614), National Institutes for Health, Washington, D.C. Data were collected by the Institute for Survey Research at Temple University, Philadelphia, PA.

such as loss of appetite and sleeplessness. Such information is not readily revealed to strangers. Thus, issues of self-presentation may influence responses when an interview delves into highly personal experience. The validity of self-reports on sensitive topics depends on the researcher's ability to establish the legitimacy of the study and to convince respondents of the complete confidentiality of the data they provide (Groves 1990). The key question in this research is whether survey modes differ on these abilities, leading to differential patterns of response to sensitive questions across modes.

Recent research suggests that sensitive topics are vulnerable to survey mode effects (Aquilino 1994; Luepker et al. 1989; Turner et al. 1992). In particular, mode differences in response anonymity may influence respondents' willingness to reveal psychologically sensitive information about themselves in a structured interview (Schwarz et al. 1991). In addition, the interviewer's ability to allay respondents' confidentiality concerns may vary by mode of communication and thus influence response tendencies to sensitive questions (Groves 1990).

### *2.1. The response anonymity hypothesis*

Self-administered questionnaires (SAQs) or answer sheets are often given to respondents to reduce self-presentation concerns during sensitive parts of a face-to-face interview. Self-administered forms allow respondents to answer threatening questions without revealing themselves to the interviewer or to others who may be present during the interview. SAQs provide greater response anonymity than interviewer-administered interviews in either the personal or telephone modes (Bradburn 1983) and should elicit a greater willingness among respondents to reveal sensitive information (Schwarz et al. 1991).

The effect of using self-administered forms in sensitive interviews has been shown in substance use surveys. Aquilino (1992) and Gfroerer and Hughes (1991) reported that self-administered forms yielded higher illicit drug use estimates than interviewer-administered telephone interviews. In a national probability sample of persons age 12 and older, Turner et al. (1992) randomly assigned respondents to either interviewer-administered or self-administered modes for answering drug and alcohol use questions (all interviews were done in person). Higher estimates of marijuana and cocaine use were obtained in the self-administered mode. Similar results were found in the 1988 wave of the National Longitudinal Survey of Youth (Schober et al. 1992): respondents given self-administered answer sheets reported more marijuana and cocaine use than respondents questioned directly by the interviewer.

There is increasing evidence that the self-administered forms have greater influence on minority than on white respondents (Aquilino 1992; Aquilino and LoSciuto 1990). In prior research with the same data set used in this article (Aquilino 1994), Black respondents reported significantly higher levels of cocaine and crack use and Hispanic respondents higher levels of marijuana use and drunkenness when SAQs were given out instead of having interviewers ask the questions directly. Fewer response anonymity effects were found for White respondents.

In this research, I examine response anonymity effects by comparing the scores of respondents who completed the CES-D on an anonymous self-administered form to

respondents interviewed face-to-face and by telephone. Two predictions are tested: (1) the self-administered mode will yield higher CES-D scores than either of the interviewer-administered modes (due to decreased influence of self-presentation concerns in the SAQ mode); and (2) the mode differentials will be greater for Black and Hispanic respondents than for White respondents.

## 2.2. *Confidentiality concerns: the social distance hypothesis*

The ability of interviewers to allay respondents' confidentiality concerns in a sensitive survey may vary by mode (Aquilino 1994; Groves 1990). The greater social distance between interviewer and respondent in the telephone mode, where nonverbal cues are lacking, may make it more difficult for interviewers to establish the study's legitimacy and convince respondents that the data they provide will be kept confidential. Studies have found greater item nonresponse to sensitive questions in telephone than in face-to-face interviews (Groves and Kahn 1979; Kormendi 1988). If the social distance hypothesis is true, the telephone mode should be more susceptible than the face-to-face mode to underreporting of sensitive or socially undesirable behavior (Aquilino 1994).

There is less consistent empirical evidence on this hypothesis than on the effects of response anonymity. Mangione et al. (1982) reported no differences between telephone and personal modes on most measures of alcohol use. In British alcohol surveys (Sykes and Collins 1988), the telephone survey yielded higher rates of drunkenness and more alcohol consumption than the face-to-face survey. In a sample restricted to the Los Angeles' SMSA, Aneshensel and her colleagues (1982a) randomly assigned respondents to a face-to-face or a telephone interview on health issues (no self-administered forms were used in the face-to-face mode). They found no overall effect of interview mode on total CES-D scores and no consistent pattern of effects on the individual items of this scale. With the same sample, these researchers also found no evidence of mode effects on self-rated health or on reports of recent illnesses, injuries or hospitalizations (Aneshensel et al. 1982b). Several studies, however, have found respondents less willing to reveal sensitive information by telephone than in face-to-face interviews. This pattern has been found in surveys measuring self-reported psychiatric symptoms and depression (Henson et al. 1978), willingness to report unlawful union campaign practices (Herman 1977), and self-reported drug and alcohol use among college students (Johnson et al. 1989). A meta-analysis of mode comparisons over four decades suggested that social desirability bias in answering sensitive questions tends to be higher in telephone than in face-to-face interviewing (de Leeuw and van der Zouwen 1988). Aquilino (1994) found the telephone mode yielded lower estimates of marijuana, cocaine, and crack use than the interviewer-administered face-to-face survey when sampling and screening methods were held constant across groups.

The prediction tested in this research is that mode of communication between interviewer and respondent influences response tendencies to sensitive questions, given that other survey design features are held constant across modes. Specifically, I hypothesized that the interviewer-administered face-to-face survey would yield higher CES-D scores than an identical interviewer-administered survey done by telephone (where social distance between interviewer and respondent is greater).

### 3. Methods

#### 3.1. Research design and samples

Systematic differences between survey modes may stem not only from interview mode, but from differences in sampling procedures and sample coverage. A problem in many studies comparing survey modes is that effects due to sampling are confounded with effects due to interview mode when telephone respondents are selected through Random Digit Dialing (RDD) and face-to-face respondents through multistage area probability sampling. This study was designed to isolate the effects of interview mode by holding sampling procedures and coverage constant across experimental groups. To accomplish this, a field experiment was designed in which respondents selected through identical sampling and screening procedures in a multistage area probability sample were randomly assigned to one of the three interview modes: face-to-face, telephone, or self-administered. Concurrent with this field experiment, a separate RDD sample of the same population was drawn and interviewed by telephone.

Both samples were drawn from the urban-suburban population of the United States. The sampling frame was restricted to adults aged 18 to 45 living in the 37 largest SMSAs in the U.S. Each SMSA included a city and its surrounding counties. A total of 168 urban and suburban counties comprised the sample. Rural counties were excluded to lower data collection costs. The age range was narrowed to younger adults because one goal of the study was to measure use of illicit drugs and usage becomes less common in later adulthood. In both samples, African-Americans and Hispanics were double-sampled. Interviews were conducted in English and Spanish. One respondent was randomly selected if more than one adult aged 18–45 resided in the household. Interviews were conducted in summer and fall of 1991.

##### 3.1.1. The area probability sample

Interviews were completed with 2,417 adults in the multistage area probability sample (a more detailed description of this sample was given in Aquilino 1994). Households were screened in person by interviewers for eligibility. All respondents in this sample were selected using identical sampling, screening, eligibility, and respondent selection procedures. Each address was pre-assigned randomly to one of three interview modes: (1) *SAQ*, an in-home interview in which self-administered answer sheets were used for sensitive questions; (2) *Face-to-face*, an in-home interview in which all questions were asked and responses recorded directly by the interviewer; and (3) *Telephone*, the interview was conducted by telephone from the interviewer's home. Professional field interviewers located throughout the country conducted the interviews in this sample. About one-third of each interviewer's assignment was done in SAQ mode, one-third in face-to-face mode, and one-third by telephone. Screening was completed in 94.3% of households in the area probability sample<sup>2</sup>; the interview success rate was 80.6% among eligible respondents,

<sup>2</sup> Response rates for the area probability sample were calculated as follows: of the 5,842 housing units in the sample, 5,511 were successfully screened (94.3%), 331 were not screened. Screening nonresponse included 176 refusals, 24 cases of a language barrier, and 131 cases not screened before the field period ended. Among screened households, 3,000 furnished an eligible respondent, of whom 2,417 were successfully interviewed (80.6%). The 583 nonrespondents included 395 refusals, 14 incapacitated respondents, 13 cases of a language barrier, and 161 not interviewed before the field period ended.

giving an overall response rate of 76%. The interview success rate was nearly identical across the three modes: 80.2%, 81.5%, and 80.1% respectively for the SAQ, face-to-face, and telephone modes.

### 3.1.2. The RDD sample

Concurrent with the multistage area probability sample described above, an RDD sample of the same 168 counties was drawn and 1,093 telephone interviews were completed with this sample. The RDD survey employed a clustering procedure (Waksberg 1978) to maximize the proportion of residential telephone numbers drawn into the sample (a detailed description of this sample is available upon request from the author). The screening interview and respondent selection procedures were identical to those used in the area probability sample, except that screening was conducted by telephone rather than in person. An in-house staff of telephone interviewers conducted the RDD survey from a central location (The Telephone Interviewing Center at Temple University's Institute for Survey Research). Response rates were typical for an RDD sample: the screening response rate was 79.5% and 74.0% of eligible respondents were interviewed, giving an overall response rate of 58.8%<sup>3</sup>.

## 3.2. Measurement

The questionnaire for the study measured the use of tobacco, alcohol, and illicit drugs, health and well-being, and demographic characteristics of the household. The 20-item CES-D was administered after the substance use segments of the questionnaire and prior to the demographic section.

Slight changes were made in the original CES-D. As designed by Radloff (1977), the CES-D has four response categories per item, ranging from "rarely or none of the time (less than one day)" to "most or all of the time (5–7 days)." In this study, respondents reported for each item "the number of days in the past week that they felt this way" (see Appendix for verbatim wording of the items used in this study). The scale for each item ranged from zero days to seven days. This response scale has been adopted in other large-scale surveys such as the NSFH (Sweet, Bumpass, and Call 1988) to elicit more detailed measurement of the persistence of symptoms over the past week. The content of one item was changed. The original CES-D item "I had crying spells" was replaced with a more gender-neutral item ("I felt unhappy") to reduce gender bias in the scale (Stommel et al. 1993).

Question wording, question order, and response categories were identical in all interview modes. No show cards were used in the SAQ or face-to-face modes to insure comparability to the telephone mode. In the SAQ mode, interviewers never saw or heard

<sup>3</sup> The RDD sample consisted of 5,888 telephone numbers, of which 3,627 were found to be working residential numbers. Among the working residential numbers, there were 742 screening nonrespondents and 2,885 successful screenings (79.5%). The RDD screening nonrespondents included 452 refusals, 45 cases with a language barrier, 221 cases where all calls resulted in no answer, answering machine, or busy signal during the entire field period, and 24 cases of other nonresponse. The screening furnished 1,477 eligible respondents, of whom 1,093 were interviewed. The 384 interview nonrespondents included 188 refusals, 9 cases with a language barrier, 3 incomplete interviews, 58 cases where all calls resulted in no answer, answering machine, or busy signal, 18 incapacitated respondents, and 108 respondents not interviewed before the end of the field period.

the respondent's answers. Interviewers read the instructions at the start of each sequence, and, if the respondent desired, read the questions aloud while the respondent completed the answer sheet. Answer sheets were sealed in an envelope in the respondent's presence upon completion of the interview.

### 3.3. *Analysis*

Case weights were computed for both samples. The weights adjust for oversampling, number of eligibles, differential selection probabilities, differential nonresponse, and the age-sex-race profile of the U.S. urban-suburban population in 1990. Case weights were used in computing the mean CES-D scores. Unweighted data were used in regression analyses testing for differences across groups (the results are the same if weighted data are used in the regressions). To preserve comparability among the groups, households without telephones ( $N = 169$ ) were excluded from the analysis. This reduced the sample size for the field experiment from 2,417 to 2,248. Missing data on the CES-D reduced the sample size for the area probability sample by 13 cases and for the RDD sample by 4 cases.<sup>4</sup> Models were fit in SUDAAN (Shah, Barnwell, and Bieler 1996) to account for the clustered nature of the samples in computing the correct standard errors of the regression coefficients.

## 4. **Results and Discussion**

### 4.1. *Demographic profiles*

There were no significant differences on any demographic characteristic among the three experimental groups in the area probability sample (Table 1). The unweighted distributions on sex, age, race-ethnicity, education, family income, employment status, marital status, and cohabitation were nearly identical for the self-administered, face-to-face, and telephone modes, as would be expected with random assignment to these groups. The demographic characteristics of the RDD sample are very similar to the groups in the area probability sample (Table 1). There is one significant demographic difference ( $p < .05$ ) between the samples. The RDD sample furnished a significantly higher proportion of Hispanic respondents (and a lower proportion of non-Hispanic whites) than the area probability sample. This difference may stem from greater availability of bilingual interviewers among the in-house staff for the RDD survey. Although about 18% of interviewers in both surveys were bilingual, more Spanish language interviews were completed in the RDD survey (9.1% of the total) than in the area survey (4.3% of the total) where bilingual interviewers were more geographically scattered. To control for this difference, race-ethnicity was included as a predictor in models comparing the experimental groups (from the area probability sample) to the RDD sample.

<sup>4</sup> Cases with missing data on 6 or more of the 20 CES-D items were excluded from the analysis (13 cases in the area probability sample, 4 cases in the RDD sample). If data were missing on 5 or fewer items, the CES-D score was summed over the answered items and then multiplied by a factor (20 divided by the number of items answered) to preserve comparability with cases with no missing data.

Table 1. Demographic characteristics of the samples: respondents age 18–45 in area probability and RDD samples (unweighted data)

	Area probability sample:			RDD sample	Test of significance <sup>1</sup>
	Self-administ. group	Face-to-Face group	Telephone group		
<i>N</i> of cases	759	749	740	1,093	–
Mean age (years)	32.0	31.8	32.0	31.4	ns
Sex:					
Male	43%	45%	43%	43%	ns
Female	57	55	57	57	
Race/ethnicity:					
Hispanic	16%	17%	15%	22%	$p < .01$
White	59	57	59	51	
Black	22	24	23	23	
Other	3	3	3	4	
Marital status:					
Married	47%	43%	46%	43%	ns
Sep/Div/Wid	15	15	15	15	
Never married	39	42	39	42	
Currently cohabiting:	6%	7%	6%	6%	ns
Education:					
Less than high school	11%	10%	12%	11%	ns
High school graduate	31	28	29	26	
Some college	20	21	20	20	
College graduate	28	28	27	31	
College student	10	12	12	12	
Work status:					
Full-time	69%	70%	71%	68%	ns
Part-time	13	11	13	13	
Unemployed	6	5	5	5	
Not in labor force	13	14	12	14	
Household income:					
under \$10,000	12%	11%	10%	10%	ns
\$10,000–29,999	26	29	26	28	
\$30,000–49,999	28	29	31	27	
\$50,000+	35	32	33	35	

<sup>1</sup> *F*-test was computed for age, Chi-square for all other variables. In the self-administered group (area probability sample), demographic questions were asked directly by the interviewer – self-administered forms were not used for this part of the interview.

#### 4.2. Mode effects on CES-D scores

Weighted CES-D scores by mode and race-ethnicity are shown in Table 2. The total CES-D score is summed over the 20 items and has a range from 0 to 140. Number of symptoms reflects the number of items respondents experienced on one or more days in the past week (range from 0 to 20). Number of persistent symptoms is the number of items experienced from five to seven days over the past week (Aneshensel et al. 1982a). The three interviewer-administered modes (face-to-face and telephone modes in the area probability

Table 2. Mean CES-D scores by interview mode and race-ethnicity (weighted means; unweighted number of cases in parentheses)

	Area probability sample			RDD sample
	Self-administered questionnaire	Face-to-Face interview	Telephone interview	Tel. interview
Total CES-D score:				
Full sample	23.1 (748)	17.1 (748)	17.6 (739)	17.2 (1,089)
Blacks	26.8 (160)	18.1 (177)	20.3 (165)	20.6 (249)
Whites	21.1 (465)	16.5 (446)	16.2 (455)	15.9 (591)
Hispanics	30.4 (116)	18.2 (123)	22.2 (112)	19.6 (237)
Number of symptoms:				
Full sample	8.6 (748)	6.4 (748)	6.6 (739)	6.1 (1,089)
Blacks	9.1 (160)	6.2 (177)	6.9 (165)	6.6 (249)
Whites	8.4 (465)	6.4 (446)	6.4 (455)	6.0 (591)
Hispanics	8.7 (116)	6.3 (123)	7.1 (112)	6.2 (237)
Number of persistent symptoms:				
Full sample	1.5 (748)	1.2 (748)	1.2 (739)	1.3 (1,089)
Blacks	1.9 (160)	1.4 (177)	1.5 (165)	1.6 (249)
Whites	1.2 (465)	1.1 (446)	1.0 (455)	1.2 (591)
Hispanics	2.7 (116)	1.1 (123)	1.9 (112)	1.6 (237)

Note: Both the area probability sample and RDD sample were drawn from the 37 largest SMSAs in the coterminous U.S., including cities and surrounding suburban counties (168 counties total). Respondents in the area probability sample were randomly assigned to interview mode after in person screening. Number of cases for the racial/ethnic groups do not sum to full sample totals due to missing data.

sample, and the RDD telephone mode) yielded nearly identical estimates of the mean CES-D score for this population. For the full sample and for each of the racial-ethnic categories, the self-administered mode furnished higher total depression scores than the interviewer-administered modes. The mode differentials appear to be greater for Blacks and Hispanics than for Whites. (The statistical significance of the mode differences and the mode-by-race interaction were tested in the multivariate regression models shown in Tables 3 and 4 and described below.) In the interviewer-administered modes, respondents indicated, on average, experiencing about 6 of the 20 symptoms; in the self-administered



mode they reported experiencing about 8 of the 20 symptoms. The number of persistent symptoms also appears to be slightly higher in the self-administered than in the interviewer-administered mode.

#### 4.3. Regression results: The field experiment

OLS regression models were used to test hypothesized mode effects for the total CES-D score. Regression models were not computed for number of symptoms or persistent symptoms since these scores are components of the total score. The models for the field experiment (area probability sample) with random assignment to mode are shown in Table 3. Because sampling, screening, and respondent selection procedures were identical

Table 3. OLS regression models for the effect of interview mode on total CES-D score: 18 to 45 year-old respondents randomly assigned to mode ( $N = 2,235$ ; unstandardized regression coefficients from SUDAAN)

Predictors	I Controls only	II Controls + interview mode	III Controls + interview mode + interactions
Intercept	36.01	33.68	34.82
Education (in years)	-1.00***	-.98***	-.97***
Female	4.66***	4.53***	4.57***
Age (in years)	-.25***	-.25***	-.26***
Race/ethnicity: (vs white/other)			
Black	2.42+	2.61*	-.08
Hispanic	3.23*	3.36*	-.42
Number of drinks in past month (log)	-.83*	-.79*	-.81*
Ever drunk in past month	3.04+	2.45	2.57+
Used illicit drugs: (vs never)			
in past month	9.37***	8.91***	8.79***
in past 12 months	8.61***	8.44***	8.38***
more than 12 months ago	3.46***	3.51***	3.46***
missing	2.26	2.09	1.99
Interview mode: (vs face-to-face)			
Self-Administered		6.65***	4.29**
Telephone		.61	-.75
Interactions:			
SAQ. $\times$ Black			5.27+
SAQ. $\times$ Hispanic			7.31**
Telephone $\times$ Black			2.87
Telephone $\times$ Hispanic			4.31
$R^2$	.078	.100	.104
Change in $R^2$	-	.022***	.004+

Note: Respondents were selected through a multistage area probability sample of the coterminous U.S. then randomly assigned to interview mode. Models were fit in SUDAAN and control for the clustered nature of the sample was conducted. Significance levels reflect two-tailed tests with unweighted data. Households without telephones were excluded from the sample.

+  $p < .10$

\*  $p < .05$

\*\*  $p < .01$

\*\*\*  $p < .001$

across the three groups, these aspects of survey design can be ruled out as explanations for mode differences in the field experiment. Although random assignment obviated the need for demographic control variables in these models, age, sex, education, and race-ethnicity were included so that interactions of these respondent characteristics with interview mode could be evaluated. Only significant interaction terms were retained in the final model.

Earlier analyses of this data set showed significant interview mode effects on measures of drug and alcohol use (Aquilino 1994). Since the CES-D was administered after the substance use portions of the interview, it was possible that mode differences in willingness to report drug and alcohol use influenced responses to subsequent items. To account for this possibility, measures of alcohol and drug use were included as controls in the regression models. The face-to-face mode is the omitted category in these models.

Model II (Table 3) shows that interview mode has a significant effect on CES-D scores. Use of self-administered answer sheets resulted in significantly higher depression scores than the interviewer-administered face-to-face mode. These results support the response

Table 4. Comparison of CES-D scores of the field experiment modes to an RDD sample: OLS regression models with 18 to 45 year-old respondents ( $N = 3,341$ ; unstandardized regression coefficients from SUDAAN)

Predictors	I Controls only	II Controls + interview mode
Intercept	37.61	35.80
Education (in years)	-1.07***	-1.03***
Female	3.61***	3.51***
Age (in years)	-.26***	-.27***
Race/ethnicity: (vs white/other)		
Black	2.40*	2.56**
Hispanic	1.75	2.10*
Number of drinks in past month (log)	-.62*	-.67*
Ever drunk in past month	3.11*	2.80*
Used illicit drugs: (vs never)		
in past month	6.69***	6.23***
in past 12 months	8.77***	8.58***
more than 12 months ago	2.77***	2.69***
missing	.69	.56
Interview mode: (vs RDD sample)		
Self-administered		6.77***
Face-to-face		.07
Telephone		.72
$R^2$	.073	.093
Change in $R^2$	-	.020***

Note: Respondents in the SAQ, face-to-face, and telephone modes were selected through a multistage area probability sample of the coterminous U.S. then randomly assigned to interview mode. The RDD mode was a separate survey of the same population interviewed by telephone. Models were fit using SUDAAN to control for the clustered nature of the samples. Significance levels reflect two-tailed tests with unweighted data. Households without telephones were excluded from the samples.

+  $p < .10$

\*  $p < .05$

\*\*  $p < .01$

\*\*\*  $p < .001$

anonymity hypothesis for collecting sensitive data. Respondents are more likely to report socially undesirable or embarrassing information when they do not have to reveal this information to the interviewer or to others who may be nearby. This finding is consistent with the mode effects research on substance use measurement which has consistently shown that self-administered forms increase levels of self-reported illicit drug use compared to interviewer-administered interviews (Aquilino 1992; 1994; Gfroerer and Hughes 1991; Turner et al. 1992). The present analysis suggests that the response privacy effect applies not only to questions on illegal behavior but to sensitive measures of psychological well-being as well.

The models in Table 3 provide no support for the social distance hypothesis. There was no difference in CES-D scores between face-to-face and telephone modes when both were interviewer-administered. The weighted estimates from these two modes are nearly identical (see Table 2). Thus, this analysis supports the findings of Aneshensel and her colleagues (1982a) of no differences between telephone and face-to-face administration of the CES-D when self-administered forms are not used.

Interaction terms of interview mode by age, sex, and education were not significant in the field experiment (data not shown). Significant interaction terms were found only for race-ethnicity by SAQ mode (Table 3, Model III). The direction of the interaction indicates that the positive effect of response privacy on CES-D scores was significantly stronger for Blacks and Hispanics than for Whites. This is consistent with the hypothesized interaction term for race-ethnicity and corroborates previous studies (Aquilino and LoSciuto 1990; Aquilino 1994) showing this same race-mode interaction for measures of substance use. The interaction terms of race-ethnicity by telephone mode were not significant. This shows that the telephone and face-to-face modes produced equivalent results not only for the sample as a whole, but across racial-ethnic groups.

The interaction of SAQ mode with race-ethnicity may be linked to minority groups' experience of discrimination. As members of oppressed groups subjected to more negative experiences with societal institutions, minority respondents may well be more guarded than white respondents in what they are willing to reveal to officials or to representatives of organizations not well known in the community. In most large-scale surveys, interviewers present themselves as representatives of academic institutions or governmental agencies. While this serves to legitimize the survey, it may also affect minorities' sensitivity to issues of self-presentation more so than it affects whites' sensitivity to such concerns.

#### *4.4. Comparison to the RDD sample*

The regression model comparing the three experimental modes to the RDD sample is shown in Table 4. This model was computed to address a potential criticism of the experimental design carried out in the area probability sample. The telephone mode in the field experiment may not adequately represent the results of a typical telephone survey. In the field experiment, the interviewer visited the household in person to conduct the screening interview, a situation that would never occur in an RDD telephone survey. Bradburn (1983) suggested that the random digit dial methodology provides greater anonymity to respondents than the face-to-face survey (at least if names are not taken), since neither

the name or address of the respondent will be known to the telephone interviewer. Thus, a true RDD design may provide greater response anonymity to respondents than did the telephone mode in the field experiment. To test this possibility, I compared the field experiment modes (all based on an area probability sample) to a concurrent RDD sample of the same population interviewed with the identical questionnaire.

The model in Table 4 suggests that the results of the field experiment for the telephone mode were not biased by the in person screening of households. The RDD sample furnishes CES-D estimates nearly identical to the telephone and face-to-face modes in the field experiment. CES-D scores in the SAQ mode were significantly higher than scores in the RDD sample and the magnitude of the SAQ effect is nearly constant across the samples. There is no evidence that results of the field experiment (shown in Table 3) would have been different if the households of telephone respondents had not been visited personally for screening. This strengthens the conclusion that use of self-administered forms affects measurement of depression with the CES-D compared with interviewer-administered modes.

## 5. Conclusions

This research suggests that response privacy makes a difference in sensitive surveys. Allowing respondents to answer survey questions without revealing themselves to the interviewer appears to reduce respondents' self-presentation concerns and makes it easier for them to report feelings or behaviors they might otherwise be hesitant to admit. Previously, this effect has been found most consistently in substance use studies. The present study shows that response anonymity influences not only the reporting of illicit behaviors such as drug use, but responses to sensitive psychological measures as well.

The results reported here replicate the findings of Aneshensel and her colleagues (1982a) of no difference in CES-D scores between the face-to-face and telephone modes of administration (both interviewer-administered modes). This study adds new information to research on the measurement of psychological well-being by demonstrating that the self-administered format yields different results than the interviewer-administered formats, and the direction of effects is toward less socially desirable responding in the SAQ format.

Item sensitivity is a critical factor in generalizing these findings beyond the CES-D. The response anonymity effects found in this study are most likely applicable to surveys that try to measure intimate details of respondents' behavior, thoughts or feelings, information not readily shared with strangers. I would not expect the same mode effects with less sensitive topics. In general, survey mode effects weaken as question content becomes less threatening to respondents (Bradburn 1983). Comparative studies involving general health issues, for example, typically report no effects of survey mode on responses (Aneshensel et al. 1982b; Groves et al. 1987; Herzog et al. 1983; Herzog and Rodgers 1988).

The findings of this research should not be interpreted to suggest that interviewer administration of the CES-D leads to invalid or unusable data. Indeed, the depression scale was originally designed for interviewer administration in face-to-face surveys and that methodology remains a legitimate option for data collection. The implication of the

findings is that adopting a self-administered format may ease respondents' self-presentation concerns when confronted with highly personal or sensitive survey questions.

## Appendix

Verbatim wording of CES-D Scale use in this research:

The following questions ask how you felt or behaved in the past week. For each statement, tell me the number of days in the past week that you had the experience. Tell me if it was zero days, 1 day, 2 days, and so forth.

- a. You were bothered by things that do not usually bother you.  
On how many days during the past week did you feel this way?
- b. You did not feel like eating, or your appetite was poor.
- c. You felt that you could not shake off the blues, even with help from family or friends.
- d. You felt that you were just as good as other people. (reverse coded)
- e. You had trouble keeping your mind on what you were doing.
- f. You felt depressed.
- g. You felt that everything you did was an effort.
- h. You felt hopeful about the future. (reverse coded)
- i. You felt your life had been a failure.
- j. You felt happy. (reverse coded)
- k. Your sleep was restless.
- l. You felt unhappy.
- m. You talked less than usual.
- n. You felt lonely.
- o. People were unfriendly.
- p. You enjoyed life. (reverse coded)
- q. You felt sad.
- r. You felt that people disliked you.
- s. You could not 'get going'.
- t. You felt fearful.

## 6. References

- Andresen, E.M., Malmgren, J., Carter, W., and Patrick, D. (1994). Screening for Depression in Well Older Adults: Evaluation of a Short Form of the CES-D. *American Journal of Preventive Medicine*, 10, 77–84.
- Aneshensel, C.S., Frerichs, R., Clark, V.S., and Yokopenic, P.A. (1982a). Measuring Depression in the Community: A Comparison of Telephone and Personal Interviews. *Public Opinion Quarterly*, 46, 110–121.
- Aneshensel, C.S., Frerichs, R., Clark, V.S., and Yokopenic, P.A. (1982b). Telephone Versus In-Person Surveys of Community Health Status. *American Journal of Public Health*, 72, 1017–1021.
- Aquilino, W.S. (1992). Telephone Versus Face-to-Face Interviewing for Household Drug Use Surveys. *International Journal of the Addictions*, 27, 71–91.

- Aquilino, W.S. (1994). Interview Mode Effects in Surveys of Drug and Alcohol Use: A Field Experiment. *Public Opinion Quarterly*, 58, 210–240.
- Aquilino, W.S. and LoSciuto, L. (1990). Effects of Interview Mode on Self Reported Drug Use. *Public Opinion Quarterly*, 54, 362–395.
- Bradburn, N. (1983). Response Effects. In *Handbook of Survey Research*, P. Rossi, J. Wright, and A. Anderson (eds.), 289–328. San Diego, CA: Academic Press.
- de Leeuw, E.D. and van der Zouwen, J. (1988). Data Quality in Telephone and Face-to-Face Surveys: A Comparative Meta-Analysis. In *Telephone Survey Methodology*, R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, and J. Waksberg, (eds.), 283–299. New York: Wiley and Sons.
- Gfroerer, J.C. and Hughes, A.L. (1991). The Feasibility of Collecting Drug Abuse Data by Telephone. *Public Health Reports*, 106, 384–393.
- Groves, R.M. (1990). Theories and Methods of Telephone Surveys. *Annual Review of Sociology*, 16, 221–240.
- Groves, R.M. and Kahn, R.L. (1979). *Surveys by Telephone: A National Comparison With Personal Interviews*. New York: Academic Press.
- Groves, R.M., Miller, P.V., and Cannell, C.F. (1987). Differences Between the Telephone and Personal Interview Data. In *National Center for Health Statistics, Owen Thornberry: An Experimental Comparison of Telephone and Health Interview Surveys*. Vital and Health Statistics, Series 2, No. 106. DHHS Pub. No. (PHS)87-1380. Public Health Service, Washington, D.C.
- Henson, R., Cannell, C.F., and Rothe, A. (1978). Effects of Interview Mode on Reporting of Moods, Symptoms, and Need for Social Approval. *The Journal of Social Psychology*, 105, 123–129.
- Herman, J.B. (1977). Mixed-mode Data Collection: Telephone and Personal Interviewing. *Journal of Applied Psychology*, 62, 399–404.
- Herzog, A.R. and Rodgers, W.L. (1988). Interviewing Older Adults: Mode Comparison Using Data From a Face-to-Face Survey and a Telephone Resurvey. *Public Opinion Quarterly*, 52, 84–99.
- Herzog, A.R., Rodgers, W.L., and Kulka, R. (1983). Interviewing Older Adults: A Comparison of Telephone and Face-to-Face Modalities. *Public Opinion Quarterly*, 47, 405–418.
- Johnson, T.P., Houglund, J., and Clayton, R. (1989). Obtaining Reports of Sensitive Behavior: A Comparison of Telephone and Face-to-Face Interviews. *Social Science Quarterly*, 70, 174–183.
- Kormendi, E. (1988). The Quality of Income Information in Telephone and Face-to-Face Surveys. In *Telephone Survey Methodology*, R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, and J. Waksberg (eds.), 341–355. New York: Wiley and Sons.
- Liang, J., Van Tran, T., Krause, N., and Markides, K.S. (1989). Generational Differences in the Structure of the CES-D Scale in Mexican Americans. *Journal of Gerontology: Social Sciences*, 44, S110–120.
- Luepker, R.V., Pallonen, U.E., Murray, D.R., and Pirie, P.L. (1989). Validity of Telephone Surveys in Assessing Cigarette Smoking in Young Adults. *American Journal of Public Health*, 79, 202–204.
- Mangione, T. W., Hingson, R., and Barrett, J. (1982). Collecting Sensitive Data: A

- Comparison of Three Survey Strategies. *Sociological Methods and Research*, 10, 337–346.
- Newmann, J.P. (1986). Gender, Life Strains, and Depression. *Journal of Health and Social Behavior*, 27, 161–178.
- Newmann, J.P. (1989). Aging and Depression. *Psychology and Aging*, 4, 150–165.
- Radloff, L.S. (1977). The CES-D Scale: A Self-report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, 1, 385–401.
- Schober, S., Fe Caces, M., Pergamit, M., and Brandon, L. (1992). Effect of Mode of Administration on Reporting of Drug Use in the National Longitudinal Survey. In *Survey Measurement of Drug Use: Methodological Studies*, C.F. Turner, J.T. Lessler, and J.C. Gfroerer (eds.), 267–276. DHHS Publication No. (ADM)92-1929. Washington, D.C.
- Schwarz, N., Strack, F., Hippler, H.J., and Bishop, G. (1991). The Impact of Administration Mode on Response Effects in Survey Measurement. *Applied Cognitive Psychology*, 5, 193–212.
- Shah, B.V., Barnwell, B.G., and Bieler, G.S. (1996). SUDAAN User's Manual, Release 7.0. Research Triangle Park, NC: Research Triangle Institute.
- Stommel, M., Given, B., Given, C., and Kalaian, H. (1993). Gender Bias in the Measurement Properties of the Center for Epidemiologic Studies Depression Scale (CES-D). *Psychiatry Research*, 49, 239–250.
- Sweet, J.A., Bumpass, L.L., and Call, V.R.A. (1988). The Design and Content of the National Survey of Families and Households. NSFH Working Paper #1, Center for Demography and Ecology, University of Wisconsin-Madison.
- Sykes, W. and Collins, M. (1988). Effects of Mode of Interview: Experiments in the U.K. In *Telephone Survey Methodology*, R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, and J. Waksberg (eds.), 301–320. New York: Wiley and Sons.
- Turner, C.F., Lessler, J., and Devore, J. (1992). Effects of Mode of Administration and Wording on Reporting of Drug Use. In *Survey Measurement of Drug Use: Methodological Studies*, C.F. Turner, J.T. Lessler, and J.C. Gfroerer (eds.), 177–220. DHHS Publication No. (ADM)92-1929. Washington, D.C.
- Waksberg, J. (1978). Sampling Methods for Random Digit Dialing. *Journal of the American Statistical Association*, 73, 40–46.

Received July 1996

Revised June 1997