COMPARING ORAL INTERVIEWING WITH SELF-ADMINISTERED COMPUTERIZED QUESTIONNAIRES
AN EXPERIMENT

LINCHIAT CHANG
JON A. KROSNICK*

Abstract   A previous field experiment conducted via national surveys showed that data collected via the Internet manifested higher concurrent and predictive validity and less random and systematic measurement error than data collected via telephone interviewing. To ascertain the extent to which these differences were attributable to mode *per se*, a laboratory experiment was conducted in which respondents were randomly assigned to answer questions either on a computer or over an intercom with an interviewer. Replicating findings from the national surveys, the laboratory experiment indicated higher concurrent validity, less survey satisficing, and less social desirability response bias in the computer mode than in the intercom mode. The mode difference in concurrent validity and non-differentiation was most pronounced among respondents with more limited cognitive skills. Taken together, these results suggest a potential inherent advantage of questionnaire self-administration on the computer over telephone administration.

As researchers are increasingly interested in conducting surveys via the Internet, it is important to understand whether shifting from oral administration of questions (in telephone or face-to-face interviews) to computer self-administered interviewing changes the answers that respondents provide. This paper reports the results of a laboratory experiment designed to assess the impact of this mode shift on survey responses.

Advance Access publication February 12, 2010

For this study, respondents were brought to the lab and randomly assigned to answer a questionnaire either self-administered on a computer or administered orally via an intercom with an interviewer. This design allowed us to assess whether the computer mode was associated with improved concurrent validity, less survey satisficing, and less social desirability response bias, as Chang and Krosnick (2009) found in a field experiment comparing these two modes.

The experimental design also allowed us to explore whether the mode effect was moderated by respondents' cognitive skills. Oral presentation might pose the greatest challenges for respondents with limited cognitive skills, because of the added burden imposed by having to hold a question and response choices in working memory while searching long-term memory and generating a judgment. Visual presentation of a question might reduce that burden on working memory, thereby helping people with limited cognitive skills the most. However, it may be that oral presentation makes question and response choice interpretation easier for people with limited reading skills than would visual presentation. If that is so, then any advantage of computer presentation might be confined to respondents high in cognitive skills and might even reverse among respondents with more limited skills. We explored these various possibilities.

We also examined whether administration time varied across modes. Respondents answering questions via computer could answer questions at whatever pace was optimal for them. But the nature of oral exchange in the absence of visual cues might lead both interviewers and respondents to accelerate the pace of questioning over an intercom beyond what would be optimal. So we thought respondents in the computer mode might complete the questionnaire more slowly than those in the intercom mode.

Methodology

RESPONDENTS

Respondents were undergraduates enrolled in introductory psychology classes at Ohio State University in spring 2001. They accessed an online database of all experiments available for participation that quarter and chose to sign up for this experiment in exchange for course credit. Only people who had resided in the United States for at least the past five years were eligible to participate. The respondents included 174 males and 158 females, most of them born between 1979 and 1982; 78% of the respondents were White, 11% were African-American, 2% were Hispanic, 6% were Asian, and the remaining 3% were of other ethnicities.

PROCEDURE

Respondents arrived at the experimental lab at scheduled times in groups of four to six and were individually randomly assigned to go alone into a

small soundproof private room containing either a computer on which to complete a self-administered questionnaire or intercom equipment. Respondents completed the questionnaire by their assigned mode and were debriefed and dismissed.

INTERVIEWERS

The interviewers were experienced research assistants who received training on how to administer the questionnaire, record answers, and manage the interview process. The procedures used for training these interviewers were those used by the Ohio State University Center for Survey Research. Following training, the interviewers practiced administering the questionnaire on the intercom. They were closely monitored during the interviewing process, and regular feedback was provided, as would be standard in any high-quality survey data collection firm.

MEASURES

The questions included many items similar to those used in Chang and Krosnick's (2009) national field experiment, including feeling thermometer ratings of political figures, approval of President Bill Clinton's job performance, perceived changes in past national conditions, expectations of future national conditions, perceptions of the 2000 presidential candidates' personality traits, the emotions evoked by the candidates, preferences on policy issues, political party identification, and liberal/conservative ideology. The measurement and coding of these variables are described in the appendix.

Respondents were asked to identify the most important problem facing the country, the most important problem facing young people in the country, the most important environmental problem facing the country, and the most important international problem facing the country. Each question offered respondents four response options. Half of the respondents (selected randomly) were offered the options in sequence A, B, C, D, whereas the other half were offered the options in sequence D, C, B, A.

Of the 332 total respondents, 205 granted permission authorizing us to obtain their verbal and math SAT or ACT test scores from the university registrar's office. All ACT scores were converted into SAT scores using the concordance table available at the College Board Web site (www.collegeboard.com), which shows the equivalent SAT scores for each corresponding ACT score. Total SAT scores were recoded to range from 0 to 1; the lowest total score of 780 was coded 0, and the highest total score of 1480 was coded 1.

Table 1. Unstandardized Regression Coefficients of Variables Predicting the Difference Between the Gore and Bush Thermometers (standard errors in parentheses)

	Intercom	Computer	*z*-test
Clinton Approval: Job	.52** (.08)	.88** (.09)	2.95**
	N = 166	N = 166	
Clinton Approval: Economy	.35** (.12)	.78** (.13)	2.35**
	N = 166	N = 166	
Clinton Approval: Foreign Relations	.27* (.11)	.79** (.12)	3.2**
	N = 166	N = 166	
Clinton Approval: Crime	.07 (.10)	.90** (.13)	5.06**
	N = 166	N = 164	
Clinton Approval: Education	.08 (.10)	.74** (.12)	4.17**
	N = 166	N = 166	
Clinton Approval: Race Relations	.22 (.12)	.84** (.13)	3.51**
	N = 166	N = 164	
Clinton Approval: Pollution	-.13 (.10)	.69** (.14)	4.67**
	N = 166	N = 165	
Clinton Approval: Health Care	.16 (.10)	.83** (.11)	4.32**
	N = 166	N = 166	
Past Conditions: Economy	.34** (.12)	.40* (.16)	.33
	N = 166	N = 164	
Past Conditions: Foreign Relations	.29* (.11)	.56** (.15)	1.46
	N = 166	N = 163	
Past Conditions: Crime	.07 (.11)	.54* (.14)	2.69**
	N = 166	N = 164	
Past Conditions: Education	.15 (.13)	.48** (.14)	1.75*
	N = 166	N = 164	
Past Conditions: Race Relations	.04 (.14)	.33* (.17)	1.35
	N = 166	N = 164	
Past Conditions: Pollution	-.10 (.12)	.55** (.14)	3.65**
	N = 166	N = 164	
Past Conditions: Health Care	.26* (.10)	.76** (.14)	2.92**
	N = 166	N = 162	
Expectations: Economy	.56** (.05)	.82** (.05)	3.63**
	N = 166	N = 164	
Expectations: Foreign Relations	.47** (.05)	.76** (.05)	3.99**
	N = 166	N = 166	
Expectations: Crime	.41** (.07)	.81** (.06)	4.34**
	N = 166	N = 166	
Expectations: Education	.52** (.06)	.79** (.05)	3.46**
	N = 166	N = 166	
Expectations: Race Relations	.54** (.09)	.82** (.07)	2.60**
	N = 166	N = 166	

Table 1. *Continued*

	Intercom	Computer	*z*-test
Expectations: Pollution	.22** (.08)	.57** (.07)	3.31**
	N = 166	N = 166	
Expectations: Health Care	.45** (.06)	.76** (.06)	3.66**
	N = 166	N = 166	
Candidates' Traits: Moral	.50** (.07)	.79** (.07)	2.98**
	N = 166	N = 166	
Candidates' Traits: Really Cares	.68** (.05)	.84** (.04)	2.37**
	N = 166	N = 166	
Candidates' Traits: Intelligent	.30** (.08)	.82** (.07)	5.08**
	N = 166	N = 166	
Candidates' Traits: Strong Leader	.57** (.06)	.76** (.05)	2.48**
	N = 166	N = 166	
Evoked Emotions: Angry	.74** (.05)	.85** (.05)	1.61
	N = 166	N = 166	
Evoked Emotions: Hopeful	.64** (.05)	.84** (.04)	3.02**
	N = 166	N = 166	
Evoked Emotions: Afraid	.70** (.09)	.83** (.07)	1.08
	N = 166	N = 166	
Evoked Emotions: Proud	.67** (.05)	.88** (.05)	2.99**
	N = 166	N = 166	
Party Identification	.77** (.09)	1.32** (.10)	4.28**
	N = 166	N = 166	
Political Ideology	.47** (.10)	.88** (.13)	2.51**
	N = 166	N = 166	
Military Spending	.30** (.10)	.45** (.14)	.87
	N = 166	N = 166	
Welfare Spending	.39** (.10)	.61** (.10)	1.51
	N = 166	N = 166	
Help for Black Americans	.53** (.12)	.65** (.12)	.71
	N = 166	N = 166	
Gun Control	.23* (.11)	.65** (.15)	2.29*
	N = 166	N = 166	
Effort to Control Crime	-.12 (.11)	.47* (.19)	2.65**
	N = 166	N = 165	
Immigration Restriction	.12 (.10)	.30* (.14)	1.02
	N = 166	N = 166	

* $p < .05$; ** $p < .01$.

Results

CONCURRENT VALIDITY

Concurrent validity of the measures was estimated using the same approach as was employed by Chang and Krosnick (2009). Table 1 displays unstandardized regression coefficients estimating the effects of 38 postulated predictors on the feeling thermometer ratings of George W. Bush subtracted from the feeling thermometer ratings of Al Gore.[1] The computer data yielded significantly higher concurrent validity than did the intercom data for 29 of these predictors. In no instance did the intercom data manifest significantly higher concurrent validity than the computer data. Across all coefficients shown in Table 1, a sign test revealed statistically significantly higher concurrent validity in the computer data than in the intercom data ($p < .001$).

To explore whether the mode difference varied in magnitude depending upon individual differences in cognitive skills, we regressed the difference in thermometer ratings on each predictor, a dummy variable representing mode, cognitive skills, and two-way interactions of mode x the predictor, cognitive skills x the predictor, and mode x cognitive skills, and the three-way interaction of mode x the predictor x cognitive skills.[2] The three-way interaction tested whether the mode effect on concurrent validity was different for people with varying levels of cognitive skills. We estimated the parameters of this equation using each of the 38 predictors listed in Table 1.

The three-way interaction was negative for 84% (32) of the predictors (seven of them statistically significant) and positive for six predictors (none statistically significant). A sign test revealed that the three-way interaction was more likely to be negative than positive ($p < .001$), indicating that the mode difference was more pronounced among respondents with limited cognitive skills. Among participants in the bottom quartile of cognitive skills ($N = 52$), the computer data yielded significantly higher concurrent validity than the intercom data for 16 out of 38 predictors, whereas among participants in the top quartile of cognitive skills ($N = 53$), the two modes did not yield statistically significantly different concurrent validity for any of the 38 predictors. Thus, it seems that respondents high in cognitive skills could manage the two modes equally well, whereas respondents with more limited cognitive skills were especially challenged by oral presentation.

---

1. Policy preferences on pollution by businesses did not predict the difference in feeling thermometer ratings regardless of mode and were therefore excluded from our concurrent validity analyses.
2. For efficiency, the massive tables showing detailed coefficients for all main effects and interaction effects are not presented here. These tables are available from the authors upon request.

SURVEY SATISFICING

Non-differentiation: Non-differentiation was measured using responses to the eight feeling thermometer questions with a formula developed by Mulligan et al. (2001). Values can range from 0 (meaning the least non-differentiation possible) to 1 (meaning the most non-differentiation possible). Intercom respondents ($M = .50$) manifested significantly more non-differentiation than the computer respondents on the feeling thermometers ($M = .44$), $t = 3.14$, $p < .01$. To test whether the mode difference in satisficing was contingent on individual differences in cognitive skills, we ran an OLS regression predicting the non-differentiation index using mode, cognitive skills, and the interaction between mode and cognitive skills. The interaction was negative and statistically significant, indicating that the mode difference in non-differentiation was more pronounced among respondents with more limited cognitive skills ($b = -.15$, $p < .05$).

Response order effects: When asked the four "most important problem" questions, half of the respondents were offered the response options in the order of A, B, C, D, whereas the other half were offered the options in the order of D, C, B, A. We computed a composite dependent variable by counting the number of times each respondent picked response option A or B, which were the first or second response option for half of the respondents and the third or fourth response option for the other half. This composite variable ranged from 0 to 4, where 0 indicates that a respondent never picked response option A or B across all four "most important problem" items, and 4 indicates that a respondent always picked response option A or B. Then, within each mode, this composite dependent variable was regressed on a dummy variable representing response choice order (coded 0 for people given order A, B, C, D and 1 for people given order D, C, B, A).

A significant recency effect emerged in the intercom mode ($b = .49$, $p < .01$), indicating that response choices were more likely to be selected if they were presented later than if they were presented earlier. In contrast, no response order effect was evident in the computer mode ($b = .07$, $p > .60$). When the composite dependent variable was regressed on the dummy variable representing response choice order, cognitive skills, and the two-way interaction between response choice order and cognitive skills, a marginally significant interaction effect emerged among respondents in the intercom mode ($b = 1.77$, $p < .10$). This interaction indicates that the mode difference was substantial among people with stronger cognitive skills (computer: $b = -.10$, n.s., $N = 57$; intercom: $b = .68$, $p < .05$, $N = 68$) and invisible among respondents with more limited cognitive skills (computer: $b = .17$, n.s., $N = 49$; intercom: $b = .21$, n.s., $N = 49$).

SOCIAL DESIRABILITY RESPONSE BIAS

Following Chang and Krosnick (2009), we explored whether social desirability response bias varied across the modes using the question asking whether the federal government should provide more or less help for African Americans. The distributions of answers from White respondents differed significantly across the two modes, $\chi^2 = 16.78$, $p < .01$. White intercom respondents were more likely than White computer respondents to say the government should provide more help to African Americans (49% in intercom mode versus 36% in computer mode), whereas White computer respondents were more likely to say the government should provide less help to African Americans (16% in intercom mode versus 38% in computer mode). This suggests that the computer respondents were more comfortable offering socially undesirable answers than were the intercom respondents.

COMPLETION TIME

One possible reason why the intercom interviews might have yielded lower response quality is the pace at which they were completed. If the lack of visual contact in intercom interactions leads interviewers and respondents to avoid awkward pauses and rush through the exchange of questions and answers, whereas self-administration allows respondents to proceed at a more leisurely pace, then the completion times for the intercom interviews may have been less than the completion times for the computer questionnaire completion.

In fact, however, the intercom interviews took significantly longer to complete than the self-administered surveys on computers, $t(330) = 21.68$, $p < .001$. Respondents took an average of 17.3 minutes to complete the self-administered questionnaire, whereas the intercom interviews lasted 26.6 minutes on average.

Discussion

Data collected via computer manifested higher concurrent validity than data collected via intercom, replicating the results of Chang and Krosnick's (2009) national survey field experiment. In addition, we found more satisficing in the intercom data than the computer data, as evidenced by more non-differentiation and a stronger response order effect. This set of evidence suggests that certain features of the computer mode may have facilitated optimal responding.

The advantage of the computer over the intercom in terms of concurrent validity and non-differentiation was especially pronounced among respondents with more limited cognitive skills and was weaker among people with stronger skills. This is consistent with the notion that the computer

may have reduced the cognitive demands imposed by oral presentation, so the greatest gap between the two modes appeared among the people most likely to be over-burdened by oral presentation. However, it is important to note that moderation of the response order effect by mode ran in the reverse direction: the computer mode manifested a significantly weaker response order effect than the intercom among respondents high in cognitive skills, whereas the mode difference was invisible among people with more limited cognitive skills. This surprising finding raises the possibility that the role of cognitive skills in moderating mode effects may be complex rather than simple. We look forward to future research investigating this issue.

Some past studies have shown that visual presentation of questions on paper yielded primacy effects, whereas oral presentation yielded recency effects (Bishop et al. 1988; Schwarz, Hippler, and Noelle-Neumann 1992). The present data replicated the expected recency effects in the intercom mode, but no response order effect appeared in the computer mode. This lack of effect in the visual mode may be due to the fact that the self-administered questionnaires were presented on computers instead of paper. Past research has shown that respondents answering questions via computer made fewer completion mistakes, left fewer items blank, and refused to answer fewer items than did paper-and-pencil respondents (Kiesler and Sproull 1986). Computer-assisted self-interviewing (CASI) has worked well even with respondents with no familiarity with computers, and respondents prefer CASI to paper and pencil (Davis and Cowles 1989; O'Reilly et al., 1994). Therefore, it is conceivable that the primacy effects often documented with paper-and-pencil surveys may be weak or non-existent in the computer mode.

Perhaps due to the absence of a human interviewer, computer respondents were apparently more willing to provide honest answers that were not socially admirable. This mode difference in social desirability bias jibes nicely with a set of past relevant findings. Respondents' reports of drinking behavior and income were more accurate in mail surveys than in face-to-face or telephone interviews (De Leeuw 1992); Catholics were more likely to endorse birth control and Jews were more likely to endorse legalized abortion on mail questionnaires than in telephone interviews (Wiseman 1972); and marital adjustment scores obtained over the telephone were higher than those obtained from mail questionnaires (Gano-Phillips and Fincham 1992). In a national follow-up survey of Medicare beneficiaries who had surgery for prostate cancer, mail respondents were more willing to report personal problems and worse health statuses than telephone respondents (Fowler, Roman, and Di 1998). Respondents were twice as likely to report unprotected sex with a non-primary partner in a mail survey than in a telephone interview, and half as likely to report volunteering in AIDS efforts (Acree et al. 1999).

In short, evidence suggests that self-administration decreases concerns with impression management, so people are less likely to conform to social desirability standards and more likely to provide honest answers to threatening or sensitive questions (Sudman and Bradburn 1974). Our evidence differs from many past studies in that random assignment to mode here means that the observed differences between modes must be due to mode effects and not to differences between the samples of people who contributed data via the two modes. The reduction in the social desirability bias in the computer mode observed here may also have partly accounted for the higher concurrent validity documented in that mode.

We hope that this experiment sets the stage for future experimental studies exploring the underlying mechanisms of the mode difference we observed. Specifically, meticulous designs are needed to investigate which features of computer self-administration account for this mode's advantage over oral interviews. The advantage could be due to the lack of standardization of oral administration across interviewers, pacing differences between modes (allowing respondents to move quickly through items they can answer easily and more slowly through items on which they need some time for reflection), reduced working-memory demands afforded by the visual presentation of questions and response options, and more. Insights into what factors are responsible for the differences we observed may shed light on possible directions for improving oral administration of survey questionnaires.

Appendix

This appendix shows the question stem and response choice wordings shown on the computer. During the oral interviews, the response options for the questions other than the feeling thermometers were preceded by "You can choose..."

Feeling thermometer ratings: "In the following list of names, please rate how favorable or unfavorable you feel toward each person by picking a number between 0 and 100. The larger the number you pick, the more you like the person. Ratings between 50 and 100 mean that you feel favorable toward the person, and ratings between 0 and 50 mean that you feel unfavorable toward the person. You would rate a person at 50 if you don't feel favorable or unfavorable. If you don't recognize a name, please enter the number 800 in the box next to that name." Rated politicians were: Bill Clinton, Al Gore, George W. Bush, Dick Cheney, Colin Powell, Jesse Jackson, Janet Reno, and John Ashcroft. For half of the respondents, the sentence "You would rate a person at 50 if you don't feel favorable or unfavorable" was not offered. All thermometer ratings were divided by 100, so that responses fell

within the range of 0 to 1, with larger numbers meaning more favorable ratings.

Approval of President Clinton's job performance: "Do you approve, disapprove, or neither approve nor disapprove of the way Bill Clinton has handled..." "His job as president," "the U.S. economy," "U.S. relations with foreign countries," "crime in America," "education in America," "relations between Black Americans and White Americans," "pollution and the environment," "health care in America." (Response options: strongly approve, approve not strongly, neither approve nor disapprove, disapprove not strongly, strongly disapprove.) For half of the respondents, the choice "neither" was not offered. Responses were coded to range from 0 to 1, with 1 indicating the most approval.

Perceived changes in past national conditions: "Next are some questions on whether you believe some things in the country now are better or worse than how they were when Bill Clinton became president in 1993, or whether each of these things is pretty much the same now as it was then. Compared to eight years ago, would you say that each of these is now much better, somewhat better, about the same, somewhat worse, or much worse?" "The nation's economy," "U.S. relations with foreign countries," "the amount of crime in America," "education in America," "relations between Black Americans and White Americans," "the amount of pollution in America," "health care in America." For half of the respondents, the choice "about the same" was not offered. Responses were coded to range from 0 to 1, with 1 indicating the most improvement over the past eight years.

Expectations of national conditions if the candidate were elected: "Now, what would you expect to happen in the country during the next four years if Al Gore was elected president in the elections last year? If Al Gore was elected president, would you expect each of the following to get better, worse, or stay the same over the next four years?" (Response choices: much better, somewhat better, about the same, somewhat worse, much worse.) "The nation's economy," "U.S. relations with foreign countries," "the amount of crime in America," "education in America," "relations between Black Americans and White Americans," "the amount of pollution in America," "health care in America."

    "Now, what would you expect to happen in the country during the next four years if George W. Bush was elected president in the elections last year? If George W. Bush was elected president, would you expect each of the following to get better, worse, or stay the same over the next four years?" (Response choices: much better, somewhat better, about the same, somewhat worse,

much worse.) "The nation's economy," "U.S. relations with foreign countries," "the amount of crime in America," "education in America," "relations between Black Americans and White Americans," "the amount of pollution in America," "health care in America." For half of the respondents, the choice "about the same" was not offered. For each issue, ratings of expectations under Bush were subtracted from expectations under Gore, and the result was coded so that it could range from 0 to 1.

Perceptions of candidates' traits: "In your opinion, how well do each of these words and phrases describe Al Gore? Extremely well, very well, somewhat, or not at all?" "Moral," "really cares about people like you," "intelligent," "can provide strong leadership."

"In your opinion, how well do each of these words and phrases describe George W. Bush? Extremely well, very well, somewhat, or not at all?" "Moral," "really cares about people like you," "intelligent," "can provide strong leadership."

For each trait, ratings for Bush were subtracted from ratings for Gore, and the result was coded so that it could range from 0 to 1.

Emotions evoked by the candidates: "When you think of Al Gore, does he make you feel..." "Angry?" "Hopeful?" "Afraid?" "Proud?" (Response options: extremely, very, somewhat, a little, not at all)

"When you think of George W. Bush, does he make you feel..." "Angry?" "Hopeful?" "Afraid?" "Proud?" (Response options: extremely, very, somewhat, a little, not at all)

For the two positive emotions, ratings for Bush were subtracted from ratings for Gore, and the result was coded so it could range from 0 to 1. For the two negative emotions, ratings for Gore were subtracted from ratings for Bush, and the result was coded so it could range from 0 to 1.

Policy preferences: "Next are a set of questions about what you think the government should do on a number of issues." "Do you think the federal government should spend more money on the military, less money on the military, or about the same amount as it spends now?" "Do you think the federal government should spend more money on social welfare programs to help the poor, less money on those programs, or about the same amount as it spends now?" "Do you think the federal government should do more to help Blacks than it does now, less to help Blacks than it does now, or about the same amount as it does now?" "Do you think the federal government should make it more difficult for people to buy a gun than it is now, make it easier for people to buy a gun, or keep these rules about the same as they are now?" "Do you think the federal government should let businesses pollute the environment more than they can now, require businesses to pollute less

than they do now, or let them pollute the environment about the same amount as they can now?" "Do you think the federal government should do more to fight crime than it does now, less to fight crime than it does now, or about the same amount as it does now?"

For each of the above questions, the offered response options were "a lot more, a little more, about the same, a little less, or a lot less." Answers were coded to range from 0 to 1.

"Do you think the federal government should make it more difficult for people from other countries to move to the United States than it is now, make it easier for them to move to the United States, or keep these rules about the same as they are now?" (Response options: a lot more difficult, a little more difficult, about the same, a little easier, a lot easier.) Answers were coded to range from 0 to 1.

For half of the respondents, the choice "about the same" was not offered in any of the above questions.

Most important problems: "Which of the following do you think is the most important problem facing this country today? Poor quality of government, drugs, rising prices, crime and violence."

"Which of the following do you think is the most important problem facing young people in this country today? Drugs and alcohol, crime and violence, pregnancy and abortion, peer pressure."

"Which of the following do you feel is the most important environmental problem facing the country today? Global warming, air pollution, saving endangered species, toxic waste disposal."

"Which of the following do you feel is the most important international problem facing the United States today? Illegal aliens, loss of respect for the United States abroad, foreign aid spending, terrorism."

Half of the respondents (selected randomly) were offered the response options in the order shown above, and the other half of the respondents saw or heard the response options in the reverse order.

Party identification: "Generally speaking, are you a Republican, Democrat, or Independent? On a 7-point scale where 1 means strong Republican, 7 means strong Democrat, and 4 means totally Independent, where would you place yourself?" Answers were coded to range from 0 (meaning strong Republican) to 1 (meaning strong Democrat).

Ideology: "When it comes to politics, would you describe yourself as liberal or conservative? On a 5-point scale where 1 means very conservative and 5 means very liberal, where would you place yourself?" Answers were coded to range from 0 (meaning very conservative) to 1 (meaning very liberal).

References

Acree, Michael, Maria Ekstrand, Thomas J. Coates, and Ron Stall. 1999. "Mode Effects in Surveys of Gay Men: A Within-Individual Comparison of Responses by Mail and by Telephone." *Journal of Sex Research* 36:67–75.

Bishop, George F., Hippler Hans-Juergen, Schwarz Norbert, and Strack Fritz. 1988. "A Comparison of Response Effects in Self-Administered and Telephone Surveys." In *Telephone Survey Methodology*, eds. Robert M. Groves, Paul P. Biemer, Lars E. Lyberg, James T. Massey, William L. Nicholls and Joseph Waksberg. New York: Wiley.

Chang, LinChiat, and Jon A. Krosnick. 2009. "National Surveys via RDD Telephone versus the Internet: Comparing Sample Representativeness and Response Quality." *Public Opinion Quarterly* 73:641–78.

Davis, Caroline, and Michael Cowles. 1989. "Automated Psychological Testing: Method of Administration, Need for Approval, and Measures of Anxiety." *Educational & Psychological Measurement* 49:311–20.

De Leeuw, Edith D. 1992. *Data Quality in Mail, Telephone and Face to Face Surveys*. Amsterdam: T.T.-publikaties.

Fowler, Floyd J., Anthony M. Roman, and Di Zhu Xiao. 1998. "Mode Effects in a Survey of Medicare Prostate Surgery Patients." *Public Opinion Quarterly* 62:29–46.

Gano-Phillips, Susan, and Frank D. Fincham. 1992. "Assessing Marriage via Telephone Interviews and Written Questionnaires: A Methodological Note." *Journal of Marriage and the Family* 54:630–5.

Kiesler, Sara, and Lee S. Sproull. 1986. "Response Effects in the Electronic Survey." *Public Opinion Quarterly* 50:402–13.

Mulligan, Kenneth, Jon A. Krosnick, Wendy Smith, Melanie Green and George Bizer. 2001. *Nondifferentiation on Attitude Rating Scales: A Test of Survey Satisficing Theory* Manuscript under review.

O'Reilly, James M., Michael L. Hubbard, Judith T. Lessler, Paul P. Biemer, and Charles F. Turner. 1994. "Audio and Video Computer Assisted Self-Interviewing: Preliminary Tests of New Technologies for Data Collection." *Journal of Official Statistics* 10:197–214.

Schwarz, Norbert, Hippler Hans-Juergen, and Elisabeth Noelle-Neumann. 1992. "A Cognitive Model of Response-Order Effects in Survey Measurement." In *Context Effects in Social and Psychological Research*, eds. Norbert Schwarz and Seymour Sudman. New York: Springer-Verlag.

Sudman, Seymour, and Norman M. Bradburn. 1974. *Response Effects in Surveys: A Review and Synthesis*. Chicago: Aldine.

Wiseman, Frederick. 1972. "Methodological Bias in Public Opinion Surveys." *Public Opinion Quarterly* 36:105–8.