

COMPARING THE QUALITY OF DATA OBTAINED BY MINIMALLY BALANCED AND FULLY BALANCED ATTITUDE QUESTIONS

ERIC M. SHAEFFER
JON A. KROSNICK
GARY E. LANGER
DANIEL M. MERKLE

Abstract When measuring attitudes with questions that offer dichotomous, mutually exclusive response options, researchers can ask “fully balanced” questions (which fully state both competing points of view) or “minimally balanced” questions (which fully state one viewpoint and only briefly acknowledge the second viewpoint). The two studies reported here investigated whether the greater efficiency of the latter approach brought with it reductions in the quality of the data obtained. Two experiments embedded in national sample surveys showed that minimally balanced and fully balanced attitude questions yielded similar distributions of responses and that responses to the two question forms were equivalent in terms of concurrent validity. These studies suggest that greater efficiency can be achieved via minimal balancing at no cost in terms of data quality.

Survey researchers have been aware for some time that balancing the wording of a question that offers dichotomous, mutually exclusive response options can alter the distribution of answers that people generate. For example, in an early demonstration, Schuman and Presser (1981) asked some respondents, “If there is a serious fuel shortage this winter, do you think there should be a law requiring people to lower the heat in their homes?” and others were asked, “If there is a serious fuel shortage this winter, do you think there should be a law requiring people to lower the heat in their homes, or do you oppose such a law?” The first item was dubbed “unbalanced” because its wording focused

ERIC SHAEFFER is a PhD candidate in the Department of Psychology at Ohio State University. JON KROSNICK is Professor of Communication, Political Science, and Psychology at Stanford University and University Fellow at Resources for the Future. GARY LANGER and DANIEL MERKLE are both with ABC News. This study was supported by a National Institute of Mental Health predoctoral fellowship awarded to the first author (NIMH grant T32-MH19728). Address correspondence to Eric Shaeffer; e-mail: shaeffer.11@osu.edu.

doi:10.1093/poq/nfi028

© The Author 2005. Published by Oxford University Press on behalf of the American Association for Public Opinion Research. All rights reserved. For permissions, please e-mail: journals.permissions@oupjournals.org.

primarily on one side of the issue: passing the proposed law. The second item is “formally balanced” because adding the words “or do you oppose such a law” gives equal billing to the alternative viewpoint. The study found that formally balancing this item led fewer people to endorse the view emphasized in the unbalanced item (29.4 percent versus 38.3 percent).

Narayan and Krosnick’s (1996) meta-analysis of a series of experiments similarly showed that formal balancing significantly reduced the proportion of respondents who endorsed the viewpoint emphasized in an unbalanced item. If a question accords undue emphasis or focus on one of various competing viewpoints, respondents may be especially likely to endorse that viewpoint. To avoid focusing respondents’ attention in a particular direction and thereby pushing their responses accordingly, question writers should presumably seek to place equal emphasis on all competing viewpoints by writing formally balanced items, which will presumably yield greater measurement validity.

But there are different ways to formally balance an item such as Schuman and Presser’s (1981) fuel shortage question. Rather than adding the phrase “or do you oppose such a law?” to the unbalanced item to create what we call a “fully balanced” question, a researcher could instead simply change the word “should” to “should or should not,” yielding the following wording: “If there is a serious fuel shortage this winter, do you think there should or should not be a law requiring people to lower the heat in their homes?” This version places equal emphasis on both competing points of view and does so more efficiently (that is, with fewer added words) than does the fully balanced version. We refer to this alternative approach as “minimal balancing” because balancing is accomplished by adding just two words.

Minimal balancing may offer some practical advantages. Fewer words mean less questionnaire administration time and therefore a lower cost of data collection and, perhaps, less respondent fatigue. Furthermore, minimal balancing may enhance the clarity of a question’s meaning (because of its simplicity) and may thereby reduce the likelihood of misunderstanding. On the other hand, the “fully balanced” format may produce better data quality.

Stanley Payne (1951) was one of the first scholars to note that possibility. He proposed that question writers should “state the negative in detail. . . . The ‘or not’ may not be enough to give the negative side a fair shake” (p. 230). He also contended, “It is probably better to state the second alternative more fully than to use the ‘or not’ phrase by itself” (p. 58). Simply adding the words “or should not” may not sufficiently counteract the salience of the idea expressed by the phrase, “a law requiring people to lower the heat in their homes.” Of course, the words “or should not” balance the preceding word “should.” But full balancing, which entails a complete and explicit statement of the alternative viewpoint (for example, “or do you think there should not be a law requiring people to lower the heat in their homes”), may more effectively communicate

to respondents that the researcher does not mean to place greater focus on one view than the other. Furthermore, the salience of the two viewpoints may be truly equal only if both viewpoints are fully expressed with an equal number of words.

Schuman and Presser (1981) included a minimally balanced question in some of their experiments: "Would you favor or oppose a law which would require a person to obtain a police permit before he could buy a gun?" These researchers treated minimal balancing as equivalent to full balancing in some experiments, although they reported no evidence regarding the interchangeability of these two approaches. However, Schuman and Presser (1981) did report two experiments that compared a minimally balanced item ("Do you think the use of marijuana should be made legal, or not?") to a question that formally balanced the two viewpoints in a heavy-handed way ("Some people think the use of marijuana should be made legal. Other people think marijuana use should not be made legal. Which do you favor?"). They found no significant difference between the distributions of the answers obtained. They tentatively concluded from these two studies that minimal balancing and formal balancing produce comparable response distributions, but they called for further investigation of this issue.

One other study on this point, by Hedges (1979), predated Schuman and Presser's (1981) study and yielded similar results. In one experiment, responses to the question, "Do you think you get enough physical exercise, or not?" were compared with responses to the question "Do you think you get enough physical exercise or too little physical exercise?" Transforming "not" into "too little physical exercise" had no impact on the proportion of affirmative answers: 73 percent and 72 percent, respectively, a nonsignificant difference. In another experiment some respondents were asked, "Would you describe yourself as overweight or not?" and others were asked, "Would you describe yourself as overweight, underweight, or is your weight about right?" In response to the first question, 33 percent of respondents said they were overweight, and 34 percent said they were overweight in response to the second question, a nonsignificant difference. Thus, Hedges's results suggest no change in the distributions of answers, but it is nonetheless possible that minimal balancing compromised the validity of answers by focusing respondents on one viewpoint.

The studies we report here were done in the spirit of Schuman and Presser's (1981) call for further research on this issue. In two experiments embedded in national random digit dial (RDD) telephone surveys, one conducted by ABC News and the other by ABC News and the *Washington Post*, we compared the distributions of answers obtained from a minimally balanced question to those obtained from a fully balanced version of the same question.

We also compared the two question forms in terms of concurrent validity. Specifically, we explored whether full balancing changed the strength of associations between the target attitude question and responses to other, conceptually

relevant criterion questions that were asked identically of all respondents.¹ Because measurement error attenuates associations between such variables, if one question form yielded stronger associations than did the other form, that would suggest the former produced more valid measurements than the latter.

Study 1

DATA

A survey by ABC News was conducted by computer-assisted telephone interviewing (CATI) June 12–16, 2002, with a national sample of 1,023 American adults. Data collection was conducted by TNS Intersearch of Horsham, PA, via TNS Express.²

MEASURES

Target Attitude Measure. Near the beginning of the interview, respondents were randomly assigned to be asked one of two versions of a question about terrorism:

Protect Rights—Minimal Balance (subsample 1): “As it conducts the war on terrorism, do you think the United States government is or is not doing enough to protect the rights of American citizens?” (Coding: yes, doing enough = 0, no, not doing enough = 1).

Protect Rights—Full Balance (subsample 2): “As it conducts the war on terrorism, do you think the United States government is doing enough to protect the rights of American citizens, or do you think the government is not doing enough to protect the rights of American citizens?” (Coding: yes, doing enough = 0, no, not doing enough = 1).

Criterion Items. From among all other questions in the questionnaire, we selected criterion items that met the following requirements: (1) the question asked about the war on terrorism, terrorist attacks, or homeland security, assuring that it was conceptually relevant to the topic of the target attitude measure; (2) the question was significantly correlated with the manipulated target question in the full sample of respondents; and (3) the question preceded the manipulated target item in the questionnaire (to ensure that the criterion question responses were not affected by the manipulation of the target question’s wording). One question met these criteria, which asked whether the United States should do more to prevent terrorist attacks (“Prevent Attacks”;

1. By “criterion,” we mean a measure that taps another construct that was expected, on theoretical grounds, to be correlated with the target item. Concurrent validity is measured by the ability of target question responses to predict responses to related criterion questions (some of which were asked before the target question, and some of which were asked after the target question).

2. Records were not available to permit calculating a response rate for this survey.

see the appendix for the question wording and coding). Respondents who said the government should do more to prevent terrorist attacks also tended to say that the government was not doing enough to protect the rights of Americans with the war on terrorism ($r = .36, p < .001, N = 937$).³

EFFECTIVENESS OF RANDOM ASSIGNMENT

As expected, subsamples 1 and 2 did not differ significantly in terms of the distributions of age ($\chi^2(5) = 8.59, p = .13, N = 936$), marital status ($\chi^2(4) = .19, p = 1.00, N = 961$), employment ($\chi^2(7) = 2.58, p = .92, N = 956$), race ($\chi^2(6) = 2.06, p = .91, N = 926$), income ($\chi^2(10) = 2.93, p = .98, N = 785$), party identification ($\chi^2(3) = 2.67, p = .45, N = 931$), education ($\chi^2(6) = 2.94, p = .82, N = 963$), gender ($\chi^2(1) = 3.00, p = .08, N = 966$), census region ($\chi^2(3) = 2.79, p = .43, N = 966$), or answers to the “Prevent Attacks” question ($\chi^2(1) = .49, p = .48$).

RESULTS

Responses to “Protect Rights.” The distributions of responses to the two forms of the “Protect Rights” question did not differ significantly from one another ($\chi^2(1) = .09, p = .77$; see rows 1 and 2 of table 1). Of the respondents who were asked the minimally balanced question, 67.4 percent said that the U.S. government was doing enough to protect the rights of citizens, compared with 66.5 percent of respondents asked the fully balanced question.

Concurrent Validity. To compare the validities of the two forms of the “Protect Rights” question, we conducted bivariate logistic regressions predicting responses to the “Prevent Attacks” question from responses to the two forms of the “Protect Rights” question. The relation was somewhat stronger when using the minimally balanced form of the “Protect Rights” question (see row 1 of table 2). To gauge the significance of the difference between these associations, we conducted another logistic regression predicting responses to the “Prevent Attacks” question using responses to the “Protect Rights” question, a dummy variable representing question form (coded 0 for people who received the minimally balanced form, and 1 for people who received the fully balanced form), and the interaction of “Protect Rights” responses and question form (see row 1 of table 3). The interaction was not significant ($b = -.29, n.s., N = 937$), suggesting that there was

3. Opponents of the Patriot Act and Bush administration policies have argued that increased efforts to prevent terrorist acts come at the expense of Americans' civil rights. However, the observed positive correlation between wanting the government to do more to prevent terrorist attacks and believing the government was not doing enough to protect rights suggests that many respondents desired both improved national security and better protected civil rights. Because respondents were not explicitly asked to choose between security and rights, it was not necessary to make a tradeoff between these competing values.

Table 1. Responses to the “Protect Rights” Question

Question Wording	“Protect Rights” Responses		Total	N
	Yes, Doing Enough	No, Not Doing Enough		
Study 1				
Minimally Balanced	67.4%	32.6%	100%	479
Fully Balanced	66.5%	33.5%	100%	487
Study 2				
Minimally Balanced	64.2%	35.8%	100%	581
Fully Balanced	63.6%	36.4%	100%	585

Table 2. Regression Coefficients Estimating the Associations of “Protect Rights” Responses with Responses to Criterion Questions

Criterion Items	“Protect Rights” Question Form		
	Minimally Balanced	Fully Balanced	Difference
Study 1			
Prevent Attacks	1.73** (463)	1.44** (474)	-.29
Study 2			
Prevent Attacks	1.48** (565)	1.24** (565)	-.23
U.S. Campaign	1.07** (418)	1.39** (402)	.32
Homeland Security	1.31** (264)	1.41** (297)	.10
Priority Campaign	.08* (305)	.07* (272)	-.01
Priority Security	.10** (271)	.06 (309)	-.04
Party Campaign	1.06** (261)	1.07** (229)	.01
Party Security	.89** (248)	1.14** (269)	.25
Country Safer	1.31** (316)	2.15** (258)	.84

NOTE.—Binary logistic regressions were conducted for all questions other than the “Priority Security” and “Priority Campaign” questions, for which ordinary least squares (OLS) regressions were conducted. Numbers of cases appear in parentheses below the regression coefficients.

* $p < .05$.

** $p < .01$.

Table 3. Regressions Testing the Difference Between Question Forms in Terms of Concurrent Validity

Criterion Items	Predictors				Cox and Snell R^2	N
	“Protect Rights” Responses	Question Form	“Protect Rights” Responses × Question Form			
Study 1						
Prevent Attacks	1.73** (.22)	-.01 (.18)	-.29 (.30)		.12	937
Study 2						
Prevent Attacks	1.48** (.19)	-.15 (.16)	-.23 (.26)		.10	1,130
U.S. Campaign	1.07** (.23)	.00 (.22)	.32 (.33)		.07	820
Homeland Security	1.31** (.28)	-.17 (.28)	.10 (.40)		.09	561
Priority Campaign	.08* (.03)	-.03 (.03)	.00 (.05)		.02	577
Priority Security	.10** (.03)	.03 (.03)	-.04 (.05)		.02	580
Party Campaign	1.06** (.29)	-.23 (.28)	.01 (.43)		.05	490
Party Security	.89** (.29)	.07 (.26)	.26 (.39)		.06	517
Country Safer	1.31** (.25)	-.68* (.28)	.84* (.40)		.14	574

NOTE.—Binary logistic regressions were conducted for all questions other than the “Priority Security” and “Priority Campaign” questions, for which ordinary least squares (OLS) regressions were conducted. Standard errors appear in parentheses below the regression coefficients.

* $p < .05$.

** $p < .01$.

no reliable difference between the two forms of the “Protect Rights” question in terms of concurrent validity.⁴

Study 2

In study 1, the criterion question used to assess concurrent validity and the target question were the first two questions in the questionnaire. Various studies

4. Effects of question wording on responses are sometimes strongest among the least-educated respondents (e.g., Narayan and Krosnick 1996) and among the oldest respondents (Knäuper 1999). To explore these possibilities, we tested three-way interactions of “Prevent Attacks” responses × question form × education or age and found neither of them to be significant.

have indicated that the number of questions a person has already answered prior to a target question can affect response quality: respondent fatigue is often greater after a person has answered a lot of prior questions, and this may reduce the quality of responses. For example, later question placement increases the likelihood that respondents will manifest acquiescence response bias (Clancy and Wachsler 1971), nondifferentiation in responding to batteries of ratings using the same scale (Herzog and Bachman 1981; Kraut, Wolfson, and Rothenberg 1975), and selection of an offered “don’t know” response option (Culpepper, Smith, and Krosnick 1992; Dickinson and Kirzner 1985; Ferber 1966; Krosnick et al. 2002). Had the target item appeared later in the study 1 questionnaire, it may have been subject to more respondent fatigue. If minimal balancing places greater emphasis on the first viewpoint, fatigued respondents may have been more likely to endorse that option than if they had received a fully balanced form of the same question.⁵ We therefore conducted a second study to test these possibilities by placing minimally and fully balanced items later in a lengthy questionnaire.

Another feature of study 1 worth noting is that the concurrent validity question and the target question were asked in succession in the questionnaire. If respondents used their answers to the first of these questions to guide their answers to the later question, this may have inflated the relation between these items and made it difficult to detect differences in validity attributable to question form. If instead many questions were asked between the target item and the criterion item in the questionnaire, the likelihood of finding a difference in concurrent validity between question forms might increase. Study 2 therefore used criterion items asked at varying distances from the target question.

Study 2 respondents were asked the “Protect Rights” question used in study 1, again with random assignment of respondents to either minimally or fully balanced forms. This question appeared near the end of a long questionnaire that addressed many topics other than terrorism. Respondents were also asked eight other questions on related topics. These criterion questions appeared at the beginning, middle, and end of the questionnaire.

DATA

An ABC News/*Washington Post* poll was conducted via computer-assisted telephone interviewing (CATI) between December 12 and 15, 2002, with a representative national sample of 1,209 American adults. Data collection was conducted by TNS Intersearch of Horsham, PA. The response rate for this survey was 31 percent.⁶

5. Fatigue may also increase the likelihood of a recency effect in responses to formally balanced questions administered orally (Krosnick and Alwin 1987).

6. This is American Association for Public Opinion Research (AAPOR) response rate 3 (AAPOR 2002). We set e to .20, based on estimates from previously published RDD telephone surveys with similar field period lengths (see, for example, Keeter et al. 2000, p. 127).

MEASURES

Each respondent was asked 65 or 66 opinion questions, depending on the specific questions he or she was randomly selected to receive. The questions addressed respondents' views of President George W. Bush, various social and political issues, political parties, the prospect of military action in Iraq, terrorism, homeland security, and the Catholic Church. Half of the sample, dubbed "subsample 1," was asked the minimally balanced form of the "Protect Rights" question. The other half of the sample, dubbed "subsample 2," was asked a fully balanced version of the same question.

Eight other questions in the questionnaire met the conditions necessary to serve as a criterion in a validity analysis, asking about whether the United States was doing enough to prevent terrorist attacks ("Prevent Attacks"), Bush's handling of the U.S. campaign against terrorism ("U.S. Campaign") and homeland security ("Homeland Security"), the priority that should be given to the U.S. campaign against terrorism ("Priority Campaign") and homeland security ("Priority Security"), which political party respondents trusted most to handle the U.S. campaign against terrorism ("Party Campaign") and homeland security ("Party Security"), and whether Bush had made the country more safe and secure ("Country Safer"). All respondents were asked the first two of these questions, and some respondents (selected randomly) were asked the other six (see the appendix for question wordings and codings).

Respondents who said the government was not doing enough to protect the rights of Americans with the war on terrorism (when answering the "Protect Rights" question) were more likely to say that the United States should do more to prevent further terrorist attacks ($r = .31, p < .001$), that they disapproved of the way President Bush was handling the U.S. campaign against terrorism ($r = .27, p < .001$) and homeland security ($r = .30, p < .001$), that the president and the Congress should give the U.S. campaign against terrorism and homeland security lower priority ($r = .13, p < .001$ and $r = .15, p < .001$, respectively), that they trusted Democrats to do a better job of handling both ($r = .23, p < .001$ and $r = .23, p < .001$; for "Party Campaign" and "Party Security," respectively), and that President Bush had not made the country safer and more secure ($r = .37, p < .001$).

EFFECTIVENESS OF RANDOM ASSIGNMENT

As expected, subsamples 1 and 2 did not differ significantly in terms of the distributions of age ($\chi^2(5) = 4.99, p = .42, N = 1,151$), race ($\chi^2(2) = .42, p = .91, N = 926$), income ($\chi^2(6) = 3.78, p = .71, N = 1,059$), education ($\chi^2(5) = 1.17, p = .95, N = 1,165$), gender ($\chi^2(1) = .13, p = .72, N = 1,166$), urban/rural classification ($\chi^2(3) = 2.20, p = .53, N = 1,153$), party identification ($\chi^2(3) = 2.73, p = .44, N = 1,156$), census region ($\chi^2(3) = .91, p = .82, N = 1,166$), or

answers to the “Prevent Attacks” ($\chi^2(1) = 2.83, p = .09$), “U.S. Campaign” ($\chi^2(1) = .67, p = .42$), “Homeland Security” ($\chi^2(1) = .16, p = .69$), “Priority Campaign” ($\chi^2(3) = 4.19, p = .24$), “Priority Security” ($\chi^2(3) = 2.20, p = .53$), “Party Campaign” ($\chi^2(1) = 1.67, p = .20$), “Party Security” ($\chi^2(1) = .97, p = .33$), or “Country Safer” ($\chi^2(1) = 2.02, p = .16$) questions.

RESULTS

Responses to “Protect Rights.” The distributions of responses to the two versions of the “Protect Rights” question did not differ significantly (see rows 3 and 4 of table 1; $\chi^2(1) = .05, p = .83$): 64.2 percent and 63.6 percent of respondents said that the U.S. government was doing enough to protect the rights of citizens in response to minimally and fully balanced questions, respectively.

Concurrent Validity. To evaluate the effect of question balancing on concurrent validity, we again conducted regressions using “Protect Rights” responses to predict responses to the eight criterion questions. As table 2 displays, the fully balanced form of the target question was more strongly related than was the minimally balanced form for five of the eight criterion items (significantly for one), and the reverse pattern appeared for the remaining three criterion items. A sign test indicated that the five/three distribution was not significantly different from the four/four distribution that would be expected by chance alone ($p \leq .73$).⁷ A meta-analysis (Rosenthal and Rubin 1986) showed that the composite effect size for the question form effect combined across questions ($e_c = .00066$) was nowhere near statistically significant ($p = .69$),⁸ and a test of homogeneity of the eight effect sizes was nonsignificant ($Q = 7.44, df = 7, n.s.$), meaning that we cannot reject the hypothesis that the null meta-analytic question form effect was constant across questions (see Hedges and Pigott 2001, p. 209). This suggests that there was no reliable effect of question balancing on concurrent validity.

Discussion

In response to Schuman and Presser’s (1981) call for further research on the impact of various types of question balancing, we conducted two experiments, and our results cast doubt on Payne’s (1951) assertion that full balancing is worth the extra words it requires. Minimally balanced questions and fully

7. Three-way interactions of question wording \times “Protect Rights” responses \times education or age were again not significant.

8. Separate effect sizes were calculated using partial correlations between the interaction terms and the criterion question responses (controlling for the main effects of “Protect Rights” responses and question wording) and the degrees of freedom for the residual mean squares. Then, a composite effect size was calculated using these separate effects, the average sample size, and the typical intercorrelation among the criterion questions, which controlled for the fact that the same respondents answered multiple criterion questions.

balanced questions consistently yielded similar response distributions, and responses to these two forms of questions generally had comparable validity. Therefore, the efficiency of minimal balancing gives that approach the advantage over full balancing, because full balancing requires including many additional words in a question but appears to yield the same results.

Of course, our failure to find differences here does not mean that no future study will ever do so. We can only properly conclude that we cannot use these studies to reject the null hypothesis in the case of the one target question we examined. Therefore, future investigations should continue to explore differences between minimally balanced and fully balanced formats, using a range of different target questions to assess the generality of the results reported here. For the time being, the findings reported here suggest that economy of word usage can be accomplished at no expense in terms of response distributions and concurrent validity. We look forward to more research in the future seeking techniques to minimize cognitive burden while maintaining the accuracy of reports.

Appendix

STUDY 1

Prevent Attacks: Do you think the United States is doing all it reasonably can do to try and prevent further terrorist attacks, or do you think it should do more? (Coding: doing all it can = 0, should do more = 1).

STUDY 2

Prevent Attacks: Do you think the United States is doing all it reasonably can do to try and prevent further terrorist attacks, or do you think it should do more? (Coding: doing all it can = 0, should do more = 1).

U.S. Campaign: Do you approve or disapprove of the way Bush is handling the U.S. campaign against terrorism? (Coding: approve = 0, disapprove = 1).

Homeland Security: Do you approve or disapprove of the way Bush is handling homeland security? (Coding: approve = 0, disapprove = 1).

Priority Campaign: Please tell me what kind of priority you think Bush and the Congress should give the U.S. campaign against terrorism. (Coding: the highest priority = 0; high priority but not the highest = 1/3, middle priority = 2/3, lower priority = 1).

Priority Security: Please tell me what kind of priority you think Bush and the Congress should give homeland security. (Coding: the highest priority = 0; high priority but not the highest = 1/3, middle priority = 2/3, lower priority = 1).

Party Campaign: Which political party, the Democrats or the Republicans, do you trust to do a better job handling the U.S. campaign against terrorism? (Coding: Democrats = 1, Republicans = 0).

Party Security: Which political party, the Democrats or the Republicans, do you trust to do a better job handling homeland security? (Coding: Democrats = 1, Republicans = 0).

Country Safer: Please tell me whether the following statement applies to Bush or not . . . He has made the country safer and more secure. (Coding: yes = 0, no = 1).

References

- American Association for Public Opinion Research (AAPOR). 2002. AAPOR Outcome Rate Calculator, Version 2.0. Lenexa, KS: AAPOR.
- Clancy, Kevin J., and Robert A. Wachsler. 1971. "Positional Effects in Shared-Cost Surveys." *Public Opinion Quarterly* 35: 258–65.
- Culpepper, Irving J., Wendy R. Smith, and Jon A. Krosnick. 1992. "The Impact of Question Order on Satisficing in Surveys." Paper presented at the annual meeting of the Midwestern Psychological Association, Chicago.
- Dickinson, John R., and Eric Kirzner. 1985. "Questionnaire Item Omission as a Function of Within-Group Question Position." *Journal of Business Research* 13: 71–75.
- Ferber, Robert. 1966. "Item Nonresponse in a Consumer Survey." *Public Opinion Quarterly* 30:399–415.
- Hedges, B. M. 1979. "Question Wording Effects: Presenting One or Both Sides of a Case." *The Statistician* 28: 83–99.
- Hedges, Larry V., and Therese D. Pigott. 2001. "The Power of Statistical Tests in Meta-Analysis." *Psychological Methods* 6: 203–17.
- Herzog, A. Regula, and Jerald G. Bachman. 1981. "Effects of Questionnaire Length on Response Quality." *Public Opinion Quarterly* 45: 549–59.
- Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert M. Groves, and Stanley Presser. 2000. "Consequences of Reducing Nonresponse in a National Telephone Survey." *Public Opinion Quarterly* 64: 125–48.
- Knäuper, Bärbel. 1999. "The Impact of Age and Education on Response Order Effects in Attitude Measurement." *Public Opinion Quarterly* 63: 347–70.
- Kraut, Allen I., Alan D. Wolfson, and Alan Rothenberg. 1975. "Some Effects of Position on Opinion Survey Items." *Journal of Applied Psychology* 60: 774–76.
- Krosnick, Jon A., and Duane F. Alwin. 1987. "An Evaluation of a Cognitive Theory of Response Order Effects in Survey Measurement." *Public Opinion Quarterly* 51: 201–19.
- Krosnick, Jon A., Allyson L. Holbrook, Matthew K. Berent, Richard T. Carson, W. Michael Hanemann, Raymond J. Kopp, Robert C. Mitchell, Stanley Presser, Paul A. Ruud, V. Kerry Smith, Wendy R. Moody, Melanie C. Green, and Michael Conaway. 2002. "The Impact of 'No Opinion' Response Options On Data Quality: Non-Attitude Reduction or an Invitation to Satisfice?" *Public Opinion Quarterly* 66: 371–403.
- Narayan, Sowmya, and Jon A. Krosnick. 1996. "Education Moderates Some Response Effects in Attitude Measurement." *Public Opinion Quarterly* 60: 58–88.
- Payne, Stanley L. 1951. *The Art of Asking Questions*. Princeton, NJ: Princeton University Press.
- Rosenthal, Robert, and Donald B. Rubin. 1986. "Meta-Analytic Procedures for Combining Studies with Multiple Effect Sizes." *Psychological Bulletin* 99: 400–406.
- Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York: Academic Press.