Spring 2006

Comparing Check-All and Forced-Choice Question Formats in Web Surveys

Jolene D. Smyth
Washington State University, jsmyth2@unl.edu

Don A. Dillman
Washington State University, dillman@wsu.edu

Leah Melani Christian
Washington State University

Michael J. Stern
Washington State University, stern-michael@norc.org

Published in *Public Opinion Quarterly* 70:1 (Spring 2006), pp. 66–77; doi: 10.1093/poq/nfj007

Published online January 1, 2006.

Comparing Check-All and Forced-Choice Question Formats in Web Surveys

Jolene D. Smyth, Don A. Dillman, Leah Melani Christian,

and Michael J. Stern

Washington State University

Corresponding author – Jolene D. Smyth, email jsmyth@wsu.edu

Abstract
For survey researchers, it is common practice to use the check-all question format in Web and mail surveys but to convert to the forced-choice question format in telephone surveys. The assumption underlying this practice is that respondents will answer the two formats similarly. In this research note we report results from 16 experimental comparisons in two Web surveys and a paper survey conducted in 2002 and 2003 that test whether the check-all and forced-choice formats produce similar results. In all 16 comparisons, we find that the two question formats do not perform similarly; respondents endorse more options and take longer to answer in the forced-choice format than in the check-all format. These findings suggest that the forced-choice question format encourages deeper processing of response options and, as such, is preferable to the check-all format, which may encourage a weak satisficing response strategy. Additional analyses show that neither acquiescence bias nor item nonresponse seem to pose substantial problems for use of the forced-choice question format in Web surveys.

A common question format in Web surveys is the "check all that apply" question, for which respondents are asked to mark all that apply from among a list of options. The check-all-that-apply question format is especially compatible with Web surveys because of the availability of the HTML-box Web design feature. Rather than limiting respondents to only one answer, this feature allows multiple items to be selected, making the design of

check-all questions quite efficient. In telephone surveys, however, the check-all format is considered awkward and is seldom used. Instead, a forced-choice question format, where respondents provide an answer (e.g., yes/no) for each item in the list, is typically employed. While the forced-choice format is more efficient for telephone surveys, self-administered survey designers have avoided its use partially due to the concern that respondents will treat forced-choice questions as check-all items by marking answers only in the "yes" category and ignoring the "no" category. The ability to require responses to each item on Web surveys could override that concern; however, error messages requiring that each item to be answered may irritate respondents and cause them to terminate their participation in the survey (Best and Krueger 2004).

As a result of these tensions, it has become common practice to convert between the check-all and forced-choice formats when switching between self-administered and interview surveys and to assume that these question formats are functional equivalents (Rasinski, Mingay, and Bradburn 1994). Sudman and Bradburn (1982), however, argue that the response task in the two question formats is fundamentally different. The check-all format presents the options as a set of items from which the respondent should choose those that apply. Conversely, the forced-choice format asks respondents to provide an answer (yes or no) for each response option, a task that should encourage respondents to consider and come to a judgment about each item individually.

This difference in response task may lead respondents to use different strategies for answering when presented with check-all and forced-choice question formats. In the check-all format, for example, the task of considering the set of items and selecting those that apply may encourage respondents to avoid expending the time and effort required to answer the question optimally by choosing only the first response option(s) they can reasonably justify, a form of weak satisficing (Krosnick 1991, 1999; Krosnick and Alwin 1987). Using this strategy, respondents can quickly satisfy the requirements of the question and then proceed to the next question without giving adequate attention to the remaining response options. Such a response strategy should manifest itself in relatively fast response times, as well as in patterns of primacy where options are more likely to be selected when they appear near the top of the list than when they appear near the bottom of the list (Krosnick 1999).

In contrast, to satisfy the requirements of a forced-choice question with its explicit yes/no categories, respondents have to commit to an answer for every item. Because it encourages respondents to elaborate on and more deeply process every option, this question format should discourage a satisficing response strategy and take longer to answer (Sudman and Bradburn 1982). In addition, it should result in more options being endorsed, both because respondents process throughout the list and because they more deeply process each individual response option, making them more likely to think of reasons the options apply (Krosnick 1992; Sudman, Bradburn, and Schwarz 1996).

In addition to different response tasks, Sudman and Bradburn (1982) point out that the interpretation of the responses themselves also differs across these two question formats. For example, respondents to check-all questions may leave an option blank for a number of reasons: (1) the option does not apply to them, (2) they are neutral or undecided about it, or (3) they overlooked it. Consequently, in a check-all question we cannot conclude that

a blank option is equivalent to "does not apply." In contrast, the addition of the explicit "no" category in the forced-choice format allows for finer differentiation of the meaning of responses; options left blank can clearly be interpreted as missing. However, the explicit "no" category in forced-choice questions may have unintended consequences if respondents who are actually neutral or otherwise undecided on a particular option are more likely to agree than disagree, a form of agreeing response bias or acquiescence (Schuman and Presser 1981). Such a tendency to agree may result in respondents marking "yes" in order to avoid being disagreeable (by marking "no"), which would result in the forced-choice format artificially yielding more options marked affirmatively.

The comparability of responses from check-all and forced-choice questions has been addressed in only one published experiment of which we are aware. Rasinski, Mingay, and Bradburn (1994) compared these question formats in a mail survey field test for round three of the 1988 National Educational Longitudinal Study. Half of the respondents were assigned a version of the survey in which three questions were formatted as check-all-that-apply questions and the other half were assigned a version in which these same three questions were formatted as forced-choice questions with yes/no categories. For all three items the mean number of options marked per respondent in the forced-choice version was significantly higher than the mean number marked in the check-all version (3.03 vs. 2.86, p = .002; 2.47 vs. 1.53, p = .001; and 1.18 vs. 0.96, p = .001).

Using the results of experiments from two Web surveys and a paper survey comparison, our purpose in this research is to extend the work of Rasinski, Mingay, and Bradburn (1994) in two important ways. First, we extend their work to Web surveys by examining check-all and forced-choice question formats in the Web mode. Second, whereas the earlier study limited its analyses to behavioral and factual questions, we include both behavioral/factual (e.g., resources used at Washington State University, student group participation, and food vendors used on campus) and opinion-based questions (e.g., descriptions of the Washington State University Pullman campus, admittance criteria, and university budget adjustments) to ascertain whether the effects of switching between formats are related to the type of question being asked. In addition to these extensions, we briefly report findings related to depth of processing and satisficing, acquiescence, and item nonresponse in check-all and forced-choice questions.

Procedures

We compare check-all and forced-choice question formats using up to four experimental variations of substantively different questions from two Web surveys and one paper survey, all designed to assess the undergraduate experience at Washington State University (WSU). Design and implementation details for all three surveys are summarized in table 1. In all of the surveys students were randomly assigned to a version of the questionnaire, and all respondents received a two-dollar incentive with the survey request. Each Web respondent also received a unique identification number that he or she was required to use to access the survey. The Web surveys were designed similarly: questions appeared on their own page in black text against a colored background; answer spaces appeared in white so

as to provide contrast between the answer spaces and the background; screens were constructed with HTML tables using proportional widths to maintain the visual aspect of the screen regardless of individual users' window sizes; and font size and style were automatically adjusted using Cascading Style Sheets to accommodate various users' screen resolutions. In the paper survey questions also appeared in black text against a colored background with white answer spaces. Replicas of the questions as formatted in the studies are available in an online appendix to this article.

Table 1. Design and Implementation Details for Surveys

	Survey Date	Experimental Versions	Number of Questions	Sample Size	Completed Responses	Response Rate
Paper	Spring 2001	4	41	1,800	1,042	58%
Web	Spring 2003	4	21	3,004	1,591	53%
Web	Fall 2003	4	25	3,045	1,705	56%

Note: The response rate reported for the three studies is American Association for Public Opinion Research (AAPOR) response rate 2 (AAPOR 2004).

Findings

Results in the first three columns of table 2 unequivocally support the expectation that the forced-choice format yields more options marked affirmatively than the check-all format. Overall in the check-all formatted versions, an average of 4.1 options were marked per question. In the forced-choice versions, the average number of options marked per question was significantly higher at 5.0 ($t = -18.57$, $p = .000$). Fifteen of the sixteen comparisons were significantly different in the expected direction, and the sixteenth approached significance ($p = .054$). Moreover, 91 percent of response options were marked affirmatively more often when they appeared in the forced-choice format than when they appeared in the check-all format.

Not only did conducting the surveys via the Web allow us to extend Rasinski and colleagues' (1994) findings to a new mode, it also allowed us to collect paradata (Heerwegh and Loosveldt 2004) to examine how much time respondents spent on each question format.[1] As a result, we can begin to assess some explanations for the finding of more options being marked affirmatively in the forced-choice format. The last three columns in table 2 indicate that in all instances respondents to the forced-choice formatted questions spent significantly more time responding than did respondents to the check-all formatted questions. At minimum, respondents spent 45 percent longer on the forced-choice format, and on average they spent two and a half times longer. Some of this additional time was undoubtedly spent marking the "no" category in the forced-choice questions, a step that is not required on the check-all format; however, the magnitude of the time differences between formats suggests that respondents spent more time on the forced-choice format independent of this extra mechanical response step. These findings support the claim of Sudman and Bradburn (1982) that items are subject to deeper processing in the forced-choice format than the check-all format, and they suggest that respondents to the check-all formatted questions may be employing a satisficing response strategy.

Table 2. Comparisons between the Check-All and Forced-Choice Format for Mean Number of Options Marked Affirmatively and Mean Time (second) Spent Answering Questions

	Mean Number of Options Marked Affirmatively			Mean Time Spend Answering		
	Check-All	Forced-Choice	1-Sided *t*-test	Check-All	Forced-Choice	1-Sided *t*-test
Web Experiment #1: Spring 2003						
Q11: Resources used at WSU (10)						
Check vs. Used/Not Used	5.4	5.7	−3.41*	15.9	25.0	−13.42*
Check (R) vs. Used/Not Used (R)	5.6	6.2	−5.43*	19.2	27.9	−14.45*
Q13: Cougar varsity sports fan (15)						
Check vs. Yes/No	2.6	3.6	−4.43*	14.1	30.5	−13.92*
Check vs. Fan/Not a Fan	2.6	3.9	−5.87*	14.1	28.0	−17.73*
Q16: Student group participation (11)						
Check vs. Yes/No	1.9	2.6	−5.01*	16.5	27.1	−13.03*
Check vs. Participate/ Not Participate	1.9	2.4	−3.67*	16.5	27.0	−13.16*
Overall Mean For Survey #1	3.3	4.1	−4.96*	16.1	27.6	−9.40*
Web Experiment #2: Fall 2003						
Q3: Descriptions of campus (12)						
Check vs. Yes/No	4.4	6.6	−17.68*	11.6	35.5	−22.26*
Q6: Admittance criteria (14)						
Check vs. Yes/No	5.0	6.1	−6.27*	13.0	42.4	−20.35*
Check (R) vs. Yes/No (R)	5.2	5.9	−4.46*	13.2	44.2	−20.64*
Q11: Univ. budget adjustments (14)						
Check vs. Yes/No	3.5	4.6	−7.24*	14.4	54.5	−24.10*
Q14: Cougar varsity sports fan (15)						
Check vs. Yes/No	3.1	4.5	−6.19*	5.5	27.1	−28.02*
Q16: Food vendors on campus (9)						
Check vs. Yes/No	4.4	5.0	−3.56*	5.2	12.5	−19.57*
Check (R) vs. Yes/No (R)	4.6	5.1	−3.79*	5.6	12.6	−22.30*
Q20: Possessions in Pullman (13)						
Check vs. Yes/No	6.4	6.7	−1.94*	7.8	20.2	−19.79*
Check (R) vs. Yes/No (R)	6.6	6.9	−1.61	8.0	21.7	−22.79*
Overall Mean For Survey #2	4.8	5.7	−4.44*	9.4	30.1	−5.41*
Paper Experiment: Spring 2002						
Q5: Cougar varsity sports fan (15)						
Check vs. Yes/No	2.6	3.8	−5.94*	N/A	N/A	N/A
Overall Mean (All Surveys)	4.1	5.0	−18.57*	N/A	N/A	N/A

Note: The number of response options offered for each question is displayed in parentheses. "(R)" denotes treatments in which the options were presented in reverse order (inverted). Time outliers were removed at two standard deviations above the mean.

* $p \leq .05$

Support for this claim is bolstered by two additional findings. First, as shown in figure 1, respondents who spent over the mean response time on check-all questions marked significantly more answers on average than those who spent the mean response time or less (5.6 vs. 3.7). In fact, these respondents marked as many and often more options than all

respondents to the forced-choice questions (overall means: 5.6 vs. 5.0, respectively), suggesting that those spending more time on check-all questions were processing the response options more deeply and thus finding a greater number of response options that applied to them. In contrast, figure 2 shows that forced-choice respondents using greater than the mean response time did not mark significantly more options for most questions (15 of 19) than their counterparts who used the mean response time or less (5.2 vs. 5.0). These findings suggest that the additional time spent on the forced-choice format that we see in table 2 is sufficient for respondents to more deeply process all of the response options, such that spending even more time does not lead to more options being marked.

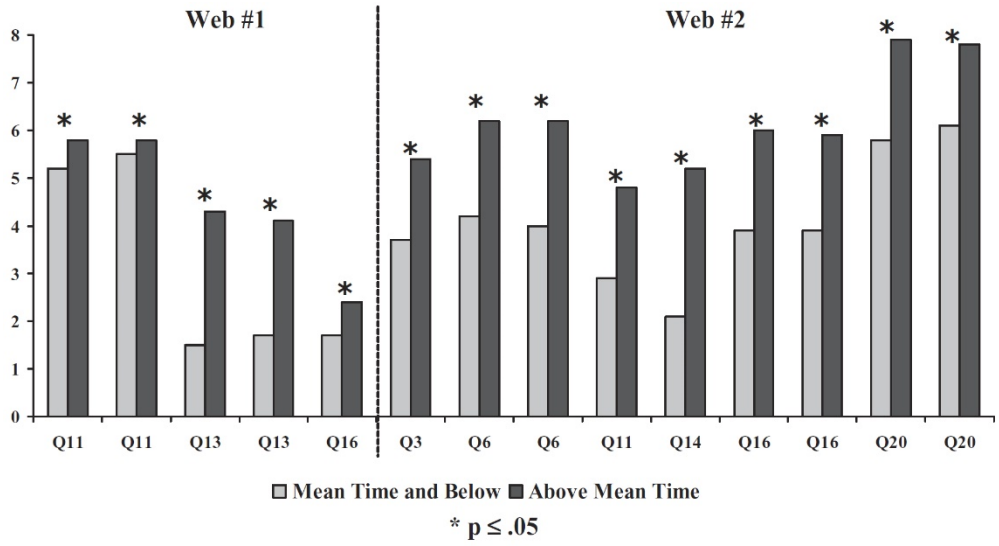

Figure 1. Mean number of options marked by those taking above and below the mean response time in the check-all format.

Second, for the check-all respondents who spent the mean response time or less, eight of ten questions presented in original and reverse order showed that options were significantly more likely to be endorsed when they appeared in the first three positions in the list than when they appeared in the last three positions (analysis not shown).[2] These patterns of primacy suggest that respondents who spend less than the mean amount of time responding to the check-all format may be employing a weak satisficing strategy. In contrast, only one such comparison resulted in significant primacy patterns for check-all respondents who spent over the mean response time.

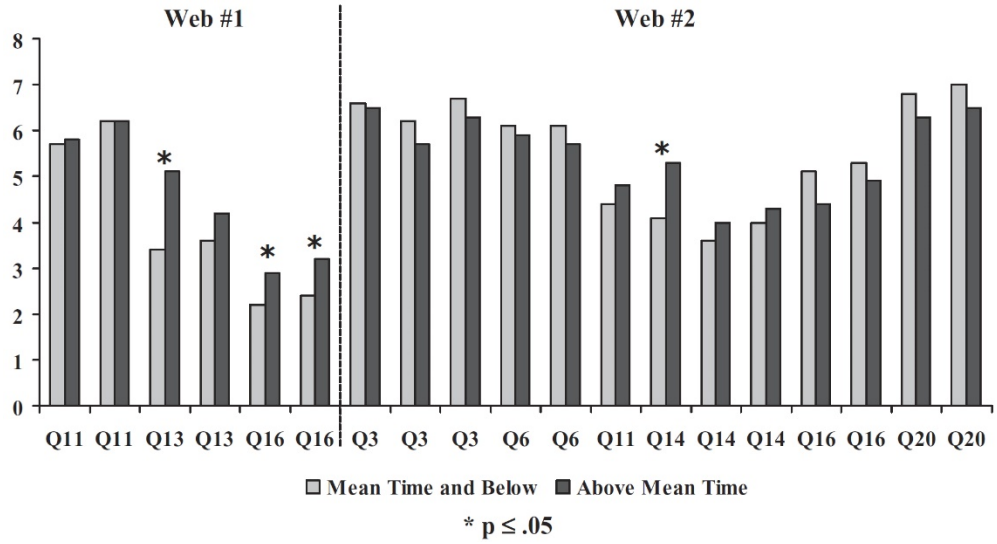

Figure 2. Mean number of options marked by those taking above and below the mean response time in the forced-choice format.

In additional analyses (not shown) we tested for acquiescence in the forced-choice format by including a third category, "don't know" or "neutral," with the yes/no categories for two questions (descriptions of WSU Pullman campus and Cougar varsity sports fan).[3] If neutral or undecided respondents are acquiescing by choosing "yes" to avoid being disagreeable, we would expect to see the third category drawing responses from the "yes" category when we compare the yes/no/don't know format to the original yes/no format. The addition of the third category did not, however, draw responses from the "yes" category for either question. In fact, for the Cougar varsity sports fan question, the "neutral" category drew responses predominantly from the "no" category. These findings indicate that "neutral" or "don't know" respondents did not choose the "yes" category in an effort to avoid rejecting items.

Finally, very few respondents treated forced-choice formatted questions as check-all questions by ignoring the "no" category and marking only within the "yes" category. Across all 24 forced-choice treatments included in the three surveys, the mean percentage of respondents who treated forced-choice questions as check-all questions was only 2.7. However, because two of the questions did have high percentages (up to 11.3 percent) of respondents using this response strategy, we investigated what made these particular questions more likely to produce check-all response patterns. We hypothesized that forced-choice questions based on opinions discourage the treatment of forced-choice questions as check-all questions because respondents are unlikely to have preformed judgments readily available to answer them and, therefore, will need extra time to form a judgment (Sudman, Bradburn, and Schwarz 1996). Thus, opinion-based questions require more consideration, which will slow the respondent down. In contrast, respondents are

more likely to have information readily available to answer behavior and fact-based questions. As a result, these questions may facilitate "quick clicking," resulting in a higher likelihood of respondents ignoring the "no" category.

The mean percentage of respondents who treated the forced-choice questions as check-all questions is 3.47 for the behavior/fact-based questions and only 1.58 for the opinion-based questions (one-sided $t = -1.55$, $p = .067$), which suggests some support for this explanation. In addition, the behavior/fact-based questions took, on average, 23.15 seconds to complete, while the opinion-based questions took 42.16 seconds to complete (one-sided $t = 4.79$, $p = .002$). Together, these findings suggest that compared with the behavior/fact-based questions, respondents gave more consideration to (or at least took longer to process) the opinion-based questions, which may have discouraged their treatment of them as check-all questions. Two points should be noted for this analysis. First, the wording of the questions included the positive and negative categories as part of the question stem (e.g., "Do you think that each description does or does not describe this campus?") to avoid prose that would encourage respondents to mark only "yes" answers (e.g., "Please check which of these sports you are a fan of"). We cannot speak to the effect of the forced-choice format on item nonresponse for questions that do not use this approach, but we think that including the positive and negative categories in the question stem is a generally advisable technique. Second, the time data should be interpreted with caution because there are substantial differences in the length of the question stems and the number and length of response options across these two types of categories that may have increased reading and comprehension time.

Discussion and Conclusions

Consistent with experimental results from a mail self-administered survey reported by Rasinski, Mingay, and Bradburn (1994) our test of ten items in two Web surveys and a paper comparison uniformly support the hypothesis that the forced-choice format results in more options being selected. Our results included item-order reversals, items with varying numbers of response options (ranging from 9 to 15), replication of one item across all three surveys, and opinion as well as behavioral items. Together with previous findings, these data strongly suggest that when self-administered surveys present respondents with the forced-choice format instead of the check-all format, respondents will select a greater number of options, regardless of question type.

Additional analyses suggested that the forced-choice format, as proposed by Sudman and Bradburn (1982), does lead respondents to more deeply process the response options, whereas a large portion of respondents to the check-all formatted questions appear to be spending less time and may not be processing all of the response options. Overall, the forced-choice respondents spent significantly more time responding to the questions, and among these respondents there was no difference in the number of options marked affirmatively by response time. In contrast, respondents who answered check-all questions quickly marked significantly fewer options and appear to have employed a weak satisficing response strategy (as evidenced by patterns of primacy), more so than their counterparts who answered these questions more slowly. Taken together, these findings support the

explanation that the increase in the mean number of response options marked in the forced-choice format compared with the check-all format is the result of deeper processing. In addition, they suggest that there is some level of "optimal" processing that respondents to the forced-choice format and those using over the mean amount of time in the check-all format are more likely to reach than those processing the check-all questions quickly. These findings raise concerns about the use of the check-all format because on average 66 percent of check-all respondents spent at or below the mean response time and, therefore, may not have reached that "optimal" processing level.

It appears that the use of the forced-choice question format, by virtue of the fact that it asks for consideration of every response option, is a desirable alternative to the use of the check-all question format for multiple-answer questions in Web surveys. The forced-choice format seems to promote deeper processing and allows for finer differentiation of meaning because options are explicitly marked negatively, but it does not encourage acquiescence, and it is not prone to high item nonresponse. Although the evidence that the forced-choice format produces "better" (Sudman and Bradburn 1982, p. 168) and more accurate responses is increasing, like Rasinski, Mingay, and Bradburn (1994) we lack external validation checks for our data and therefore cannot say with certainty that the forced-choice format produces more accurate responses. As such, this is an issue in need of further research.

In addition to external validation checks, an important next step in this research is to compare the use of the forced-choice format in aural (e.g., telephone) surveys with its use in visual, self-administered surveys. Although we do not yet know how the check-all and forced-choice question formats perform across modes, the evidence reported here from self-administered surveys clearly suggests that the forced-choice and check-all formats are not functional equivalents. These findings give ample reason to be concerned about the common practice of automatically converting between check-all and forced-choice formats when switching between self-administered and aural modes and about combining data across these two formats in mixed-mode surveys.

Acknowledgments – This research note is a revised version of a paper presented at the 2005 American Association for Public Opinion Research meeting, May 12–15, Miami Beach, Florida. Analysis of these data was supported by funds provided to the Washington State University Social and Economic Sciences Research Center under Cooperative Agreement #43-3AEU-1-80055 with the U.S. Department of Agriculture National Agricultural Statistics Service, supported by the National Science Foundation, Division of Science Resource Statistics. Data collection was financed by the Social and Economic Sciences Research Center and the Gallup Organization. A more detailed analysis of the data analyzed here, including exact question formats and tabular comparisons referenced in this article, is available from a report to the National Science Foundation by the same authors: "Comparing Check-All and Forced-Choice Question Formats in Web Surveys: The Role of Satisficing, Depth of Processing, and Acquiescence in Explaining Differences," Social and Economic Sciences Research Center Technical Report 05-029, available online at http://survey.sesrc.wsu.edu/dillman/papers.htm (accessed December 31, 2005).

Notes

1. The paradata were collected slightly differently in the two Web surveys. Specifically, in the first Web survey the time is measured from when the page loaded to when the respondent clicked their last response. Response time in the second survey is measured from when the page loaded to when the respondent clicked the "submit" button. Comparisons within surveys should not be affected by this programming difference.
2. Tables for all "not shown" analyses are available in Smyth et al. (2005).
3. Respondents who marked all of the options "yes" were excluded from these percentages as we assume they sincerely meant "yes" on all options and were not treating the question as a check-all. An additional question, Q24, is included in this analysis that is not included in previous analyses. There is no check-all treatment for this question, which precludes its inclusion in previous analyses, but that limitation is not relevant for the current analysis.

References

American Association for Public Opinion Research (AAPOR). 2004. *Standard Definitions: Final Disposition of Case Codes and Outcome Rates for Surveys*. 3d ed. Lenexa, KS: AAPOR.

Best, Samuel J., and Brian Krueger. 2004. *Internet Data Collection*. Thousand Oaks, CA: Sage.

Heerwegh, Dirk, and Geert Loosveldt. 2002. "Describing Response Behavior in Web Surveys Using Client Side Paradata." Paper presented at the International Workshop on Web Surveys, Mannheim, Germany.

Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5: 213–36.

Krosnick, Jon A. 1992. "The Impact of Cognitive Sophistication and Attitude Importance on Response-Order and Question-Order Effects." In *Context Effects in Social and Psychological Research*, ed. Norbert Schwarz and Seymour Sudman, pp. 203–18. New York: Springer-Verlag.

Krosnick, Jon A. 1999. "Survey Research." *Annual Review of Psychology* 50: 537–67.

Krosnick, Jon A., and D. F. Alwin. 1987. "An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement." *Public Opinion Quarterly* 51: 201–19.

Rasinski, Kenneth A., David Mingay, and Norman M. Bradburn. 1994. "Do Respondents Really 'Mark All That Apply' on Self-Administered Questions?" *Public Opinion Quarterly* 58: 400–408.

Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys Experiments on Question Form, Wording, and Context*. New York: Academic Press.

Smyth, Jolene D., Don A. Dillman, Leah Melani Christian, and Michael J. Stern. 2005. "Comparing Check-All and Forced-Choice Question Formats in Web Surveys: The Role of Satisficing, Depth of Processing, and Acquiescence in Explaining Differences." Social and Economic Sciences Research Center Technical Report #05-029, Washington State University. Available online at http://survey.sesrc.wsu.edu/dillman/papers.htm (accessed December 31, 2005).

Sudman, Seymour, and Norman M. Bradburn. 1982. *Asking Questions.* San Francisco: Jossey-Bass.

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. 1996. *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology.* San Francisco: Jossey-Bass.