

DOES “YES OR NO” ON THE TELEPHONE MEAN THE SAME AS “CHECK-ALL-THAT-APPLY” ON THE WEB?

JOLENE D. SMYTH
LEAH MELANI CHRISTIAN
DON A. DILLMAN

Abstract Recent experimental research has shown that respondents to forced-choice questions endorse significantly more options than respondents to check-all questions. This research has challenged the common assumption that these two question formats can be used interchangeably but has been limited to comparisons within a single survey mode. In this paper we use data from a 2004 random sample survey of university students to compare the forced-choice and check-all question formats across web self-administered and telephone interviewer-administered surveys as they are commonly used in survey practice. We find that the within-mode question format effects revealed by previous research and reaffirmed in the current study appear to persist across modes as well; the telephone forced-choice format produces higher endorsement than the web check-all format. These results provide further support for the argument that the check-all and forced-choice question formats do not produce comparable results and are not interchangeable formats. Additional comparisons show that the forced-choice format performs similarly across telephone and web modes.

JOLENE D. SMYTH is with the Survey Research and Methodology Program and the Department of Sociology at the University of Nebraska-Lincoln, PO Box 880241, Lincoln, NE 68588, USA. LEAH MELANI CHRISTIAN is with Pew Research Center for the People & the Press, 1615 L Street NW, Suite 700, Washington, DC 20036-5610, USA. DON A. DILLMAN is with Washington State University, Wilson Hall 133, PO Box 644014, Pullman, WA 99164, USA. This is a revision of a paper presented at the Second International Conference on Telephone Survey Methodology, Miami, FL, USA, January 11–15, 2006. Analysis of these data was supported by funds provided to the Washington State University Social and Economic Sciences Research Center (SESRC) under Cooperative Agreement 43-3AEU-1-80055 with the USDA-National Agricultural Statistics Service, supported by the National Science Foundation, Division of Science Resource Statistics. Data collection was financed by the SESRC and additional funds provided to the SESRC by the Gallup Organization. Address correspondence to Jolene D. Smyth; e-mail: jsmyth2@unl.edu

Introduction

It is common practice to ask multiple-response questions as a series of yes/no (forced-choice) items for telephone surveys, but to ask the same questions in a check-all-that-apply format in both mail and web surveys. However, a growing body of research indicates that the forced-choice question format tends to result in endorsement of significantly more response options than does the check-all question format (Rasinski, Mingay, and Bradburn 1994; Smyth et al. 2006; Thomas and Klein 2006). One limitation, however, of most research on this issue to date is that the comparisons of these question formats have been conducted *within* as opposed to *across* survey modes. As a result, these studies give us a greater understanding of pure question format effects, but fall short of addressing the fundamental issue of whether or not converting between-question formats to accommodate survey modes yields data of comparable quality. In other words, they cannot address the issue of interactions between question format *and* survey mode.

In this paper, we attempt to isolate question format effects from survey mode effects using data collected from a web and telephone mixed-mode survey of a random sample of Washington State University (WSU) undergraduate students in the Fall of 2004. In addition to summarizing question format effects for four questions within the web mode, we extend previous research by experimentally examining the extent to which the check-all format administered on the web produces comparable responses to the forced-choice format administered via telephone (i.e., a format comparison *across* modes). We also extend previous research by examining the effects of survey mode independent of question format effects in five forced-choice questions administered by both web and telephone modes.

Background

Three previous studies have experimentally examined the comparability of responses between the forced-choice and check-all question formats. Rasinski, Mingay, and Bradburn (1994) reported that for three experimental items in a paper questionnaire, the mean number of options endorsed per respondent was significantly greater when they were formatted as forced-choice questions than as check-all questions. Similarly, Smyth et al. (2006) reported that the forced-choice question format led to significantly more options being endorsed than the check-all format in 15 of 16 experimental comparisons involving both fact/behavior and opinion/attitude questions from two web surveys and a paper survey. Finally, Thomas and Klein (2006) reported that the forced-choice format produced significantly higher endorsement than the check-all format in five separate web experiments and did so regardless of topic, language, or country.

Smyth et al. (2006) assert that the root source of the differences between these two formats is the fundamental difference in response task. The check-all

question format allows for a weak satisficing response strategy (Rasinski, Mingay, and Bradburn 1994) whereby respondents can satisfy the requirements of the question by choosing the first options they can reasonably justify and then move on without giving ample consideration to remaining response options (Krosnick 1991, 1999; Krosnick and Alwin 1987). In contrast, the forced-choice format requires respondents to report a judgment about every response option to satisfy the requirements of the question which encourages deeper processing and discourages weak satisficing response strategies (Sudman and Bradburn 1982). In support of these assertions Smyth et al. (2006) show that (1) respondents spend significantly less time answering check-all questions; (2) consistent with weak satisficing, respondents who spend less than the mean response time answering check-all questions are more likely to endorse items when they appear on the top of the list; and (3) respondents who spend over the mean response time on check-all questions endorse as many or more options as forced-choice respondents. In contrast, they found that respondents spent significantly longer on the forced-choice formatted questions and endorsed the same number of options regardless of response time, suggesting that all respondents more deeply process the response options in this question format.

Together these studies shed considerable light on forced-choice and check-all question format differences and the processes that underlie them; however, all are limited to comparisons within paper and web self-administered modes. But few, if any, survey designers use these two question formats interchangeably in this way (i.e., within modes). The high demands that the check-all format places on respondents' memory make it an impractical question format to administer aurally, so telephone surveyors commonly pose such multiple-response questions as a series of yes/no items to reduce memory burden. In contrast, memory limitations are of less concern in web and mail surveys where information can be presented to respondents visually. Under these conditions the check-all format is often used because it is more space efficient to present and is presumed to be easier and less time consuming to complete than the forced-choice format. Thus, the practical question for most surveyors is: Do we get comparable data when we use these two question formats across modes, as is customary?

Several factors suggest that survey mode may affect responses to check-all and forced-choice questions. First, telephone interviews tend to be conducted relatively quickly (Krosnick and Alwin 1987) and require higher demands on respondents' memory—two factors that encourage preformed responses and shallower processing of response options (Schwarz et al. 1991; Dillman et al. 1996). In addition, the presence of an interviewer may result in social desirability, acquiescence, and lack of privacy to disclose sensitive information (de Leeuw 1992, 2005). We know of only two studies in which the check-all and forced-choice formats have been compared across survey modes. Jordan, Marcus, and Reeder (1980) found that the telephone forced-choice respondents endorsed more options than face-to-face check-all (on a show card) respondents.

They attributed the difference to mode effects rather than question format effects. Similarly, Nicolaas and Tipping (2006) found that forced-choice questions in a mail survey yielded higher endorsement rates than parallel check-all questions presented on show cards in face-to-face surveys. Both of these studies show that the forced-choice format produced higher endorsement regardless of survey mode; however, question format and survey mode are confounded so we are unable to determine whether the differences the authors observe are due to mode effects, question format effects, or some combination of both.

Procedures and Analytic Strategy

The data used in this paper were collected in Fall 2004 via simultaneous telephone and web surveys of a random sample of 3,408 full-time, Pullman campus WSU undergraduates. Students were randomly assigned to one of six questionnaire versions (three web and three telephone), all of which contained the same 25 questions about the WSU student experience arranged in the same order. Example screenshots from the web survey and telephone question wording can be seen in figure 1. Complete question wording is available in the Online Appendix. An attempt was made to standardize multiple-answer question formats within versions such that respondents to versions 1 and 2 on the web received only forced-choice formatted questions and, with the exception of one question, respondents to version 3 received only check-all formatted questions. All telephone respondents received only forced-choice formatted questions. To help standardize the visual stimulus across various hardware and software configurations, the web pages were designed with HTML tables using proportional widths and with cascading style sheets to adjust for font size and style.

Respondents were initially contacted via postal mail and given a \$2 incentive. Web respondents were also provided a personal identification number required to access the survey, and the approximately 2/3 for whom we had an e-mail address were sent an e-mail including both their PIN and a link to the survey. Reminders to web nonrespondents were sent via both postal mail and e-mail. Telephone respondents were contacted in up to 10 call-back attempts by the WSU Social and Economic Science Research Center's telephone lab. Response rates were 59 percent (1,054/1,800) for the web and 59 percent (945/1,608) for the telephone.¹

The experimental design we use here prohibits us from *statistically* modeling format and mode interaction effects because we are missing the necessary fourth treatment—the check-all format administered via the telephone. We chose to forego this treatment because of the impracticality of administering a check-all question over the telephone. But, while we cannot statistically

1. AAPOR (2004) Response Rate 1. There is no significant difference in start date at WSU between the web and telephone respondents.

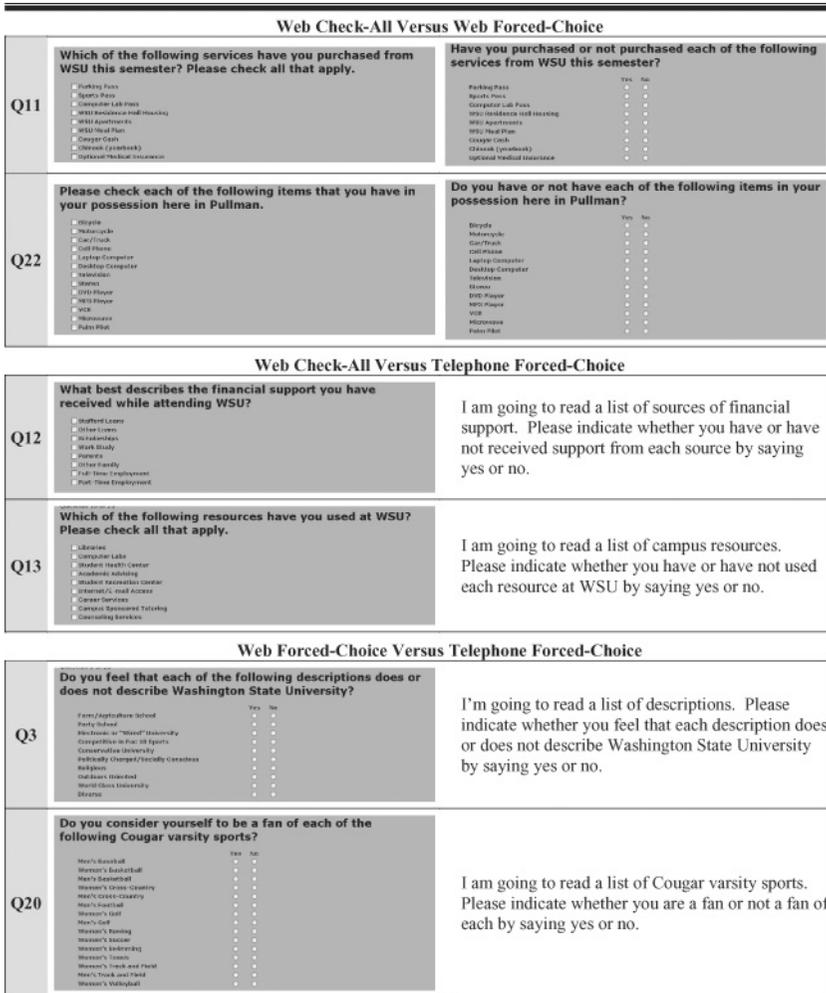


Figure 1. Examples of Web and Telephone Experimental Treatments.

test for interaction effects, the tripartite experimental design does allow us to independently address each type of effect (question format only, question format and mode, and mode only) and to look at them in concert to help shed light on how these effects may work. We start our analyses by comparing forced-choice and check-all responses *within the web mode* to determine whether the current data replicate previous findings of question format effects. We then compare the check-all format on the web to the forced-choice format on the

Table 1. Mean Number and Percentage of Items Endorsed in the Web Check-All and Forced-Choice Formats

	Forced-choice		Check-all		Difference		One-sided <i>t</i> -test	
	No.	%	No.	%	No.	%	<i>t</i>	<i>p</i>
Q11: Services purchased (9)	2.57	28.56	2.44	27.11	0.13	1.45	1.05	.147
Q13: Use of resources (9)	5.19	57.67	4.97	55.22	0.21	2.45	1.84	.033
Q20: Varsity sports fan (15)	4.71	31.40	3.11	20.73	1.60	10.67	6.06	.000
Q22: Possessions (13)	6.68	51.38	6.52	50.15	0.16	1.23	0.91	.182
Overall means	4.74	42.25	4.19	38.30	0.55	3.95	5.10	.000

NOTE.—Parentheses contain the number of response options accompanying the question. Forced-choice $n = 345$; check all $n = 342$; degrees of freedom = 685.

telephone to examine the effects of question format on responses *across modes*. This comparison addresses the fundamental question of this paper. Finally, to examine mode effects in forced-choice questions, we compare forced-choice results across telephone and web modes. Since the survey design consisted of three experimental versions of each of the web and telephone mode surveys, we are able to examine the same four questions (Q11, 13, 20, 22) across all three layers of the analysis.

Findings

Table 1 shows that in every comparison *within the web mode* the forced-choice question format yielded higher endorsement of options than the check-all format. Overall, the forced-choice format yielded an average of 4.74 of the options (42.3 percent of them) endorsed and the check-all format yielded an average of 4.19 (38.3 percent) ($t = 5.10, p \leq .000$). Individually the differences are significant in two of the four questions (Q13 and 20). Inspection of the 46 individual response options in these questions indicates that 78 percent of them were marked more often in the forced-choice format, 37 percent significantly so.² None were marked significantly more often in the check-all format.

Table 2 compares the mean number of options endorsed in the forced-choice format on the telephone and the check-all format on the web. Again, the forced-choice format consistently yields higher endorsement of options than does the check-all format with three of the five comparisons reaching significance (Q11, 12, and 20). Overall, the telephone forced-choice format yielded an average of 4.44 (41.3 percent) options endorsed while the web check-all format only yielded an average of 3.87 (37.2 percent) ($t = 6.39, p \leq .000$). Individually, 78 percent of the 54 options were marked more often in the telephone forced-choice

2. Analysis not shown. All analyses that are not shown are available upon request from the authors.

Table 2. Mean Number and Percentage of Options Endorsed in the Telephone Forced-Choice Format Compared to the Web Check-All Format

	Telephone		Web		Difference		One-sided	
	forced-choice		check-all				t-test	
	No.	%	No.	%	No.	%	t	p
Q11: Services purchased (9)	2.66	29.56	2.44	27.11	0.22	2.44	1.73	.042
Q12: Financial resources (8)	3.03	37.88	2.62	32.75	0.41	5.13	3.86	.000
Q13: Use of resources (9)	5.01	55.67	4.97	55.22	0.04	0.44	0.31	.378
Q20: Varsity sports fan (15)	5.01	33.40	3.11	20.73	1.90	12.67	7.58	.000
Q22: Possessions (13)	6.51	50.08	6.52	50.15	0.01	-0.08	-0.07	.528
Overall means	4.44	41.31	3.87	37.19	0.57	4.12	6.39	.000

NOTE.—Parentheses contain the number of response options accompanying the question. Forced-choice $n = 311$; check-all $n = 342$; degrees of freedom = 651.

format, 35 percent significantly so, and none were marked significantly more often in the web check-all format.^{3,4}

Table 3 addresses the extent to which the forced-choice question format is prone to mode effects across the telephone and web modes. On average, the telephone mode yielded 5.54 options marked (51.2 percent) and the web yielded a comparable 5.50 options marked (50.8 percent) ($t = 0.52, p \leq .606$). In seven of nine individual comparisons there is no significant difference in the mean number of options endorsed across these two modes. Inspection of the 101 individual response options shows that 48.5 percent of them were endorsed more often in the web mode (4 percent significantly) and another 48.5 percent were endorsed more often in the telephone mode (6 percent significantly). The remaining 3 percent had exactly the same endorsement rates across modes. These results indicate that the forced-choice question format performs similarly across web and telephone survey modes.⁵

3. Note that the slight variation in question wording across formats in Q12 (i.e., “the best” in the check-all format) may have contributed to the differences in this question.

4. An alternative explanation for higher endorsement rate on forced-choice questions is acquiescence, or the tendency for indifferent respondents to endorse rather than reject an option. Previous research has shown that acquiescence does not explain differences between the forced-choice and check-all formats within the web mode (Smyth et al. 2005) and because the forced-choice format performs similarly across web and telephone modes in the current study, it is unlikely that telephone respondents were significantly more prone to acquiescence than web respondents.

5. One possible explanation for the different findings across modes (i.e., table 2) is that telephone forced-choice respondents may endorse more items because they do not know what additional items will be offered. In contrast, web check-all respondents can preview the entire list before choosing which items to select so they may apply more stringent criteria and feel the need to limit their yes responses to only a selection of items. However, inasmuch as the forced-choice format performs similarly across both telephone and web modes, it seems the ability to preview the items does not have a large effect on endorsement rates.

Table 3. Mean Number and Percentage of Options Endorsed in the Telephone and Web Forced-Choice Formats

	Telephone		Web		Difference		2-sided t-test		d.f.
	No.	%	No.	%	No.	%	t	p	
Q3: Descriptions (original) (10)	6.42	64.20	6.32	63.20	0.10	1.00	0.81	.416	654
Q3: Descriptions (reverse)	6.78	67.80	6.48	64.80	0.30	3.00	2.19	.029	676
Q11: Services purchased (9)	2.66	29.56	2.57	28.56	0.09	1.00	0.70	.485	654
Q13: Use of resources (original) (9)	5.01	55.67	5.19	57.67	-0.18	-2.00	-1.53	.126	654
Q13: Use of resources (reverse)	5.13	57.00	5.07	56.33	0.06	0.67	0.50	.614	676
Q20: Varsity sports fan (15)	5.01	33.40	4.71	31.40	0.30	2.00	0.96	.336	654
Q22: Possessions (possess versus have) (13)	6.51	50.08	6.68	51.38	-0.17	-1.31	-0.99	.324	654
Q22: Possessions (possess versus possess)	6.51	50.08	6.87	52.85	-0.36	-2.77	2.02	.043	676
Q22: Possessions (have versus have)	6.71	51.62	6.68	51.38	0.03	0.23	0.16	.874	666
Overall means	5.54	51.17	5.50	50.77	0.04	0.40	0.52	.606	-

NOTE.—Parentheses contain the number of response options accompanying the question. When versions are included in multiple comparisons (as in Q22) they are only counted in the overall means once.

Additional analyses not shown here indicate that item nonresponse is not a major concern in the forced-choice format. An average of 94 percent of telephone and 97 percent of web respondents provided an answer for every item and on average less than 1 percent of web respondents treated the forced-choice questions in a check-all manner (i.e., marked only in the affirmative category and not the negative). For those who left any items blank, web respondents were at liberty to simply skip items but telephone respondents had to tell the interviewer if they did not want to provide an answer and the fact that they could do so was not explicitly disclosed. As a result, we might expect higher item nonresponse on the web than on the telephone (de Leeuw 2005), but item nonresponse rates did not differ substantially by mode. In seven of the nine comparisons there is higher nonresponse on the web, but in only one of these the difference is significant. The two comparisons in which the telephone produces significantly higher item nonresponse are both for the question asking about descriptors of WSU (Q3 – compared in original and reverse order). It is unclear why this question produced higher item nonresponse among telephone respondents, but one possible explanation is that respondents were hesitant to take a position when talking to an interviewer who was clearly associated with WSU.

Discussion and Conclusions

Our examination of question format effects, survey mode effects, and the combination of both mode and format effects in this paper produces three important findings. First, the finding reported here that web respondents endorse more options in the forced-choice format than the check-all format, along with previous research showing the same effect (Rasinski, Mingay, and Bradburn 1994; Smyth et al. 2006; Thomas and Klein 2006), indicates that there are significant question format effects between check-all and forced-choice questions within web and paper modes. Second, the fact that the comparison of these question formats across the web and telephone modes mirrors their comparison within the web mode so closely indicates that this question format effect persists across modes. Our research suggests, therefore, that the differences that Jordan, Marcus, and Reeder (1980) attributed to survey mode effects may at least partially be the result of the question format change that accompanied their mode comparison. Finally, the third finding of this paper is that the forced-choice question format performs similarly in web and telephone survey modes.

Based on these three findings and previous research which suggests that the forced-choice question format encourages deeper processing and therefore more optimal response behavior (Smyth et al. 2006), we recommend using the forced-choice question format instead of the check-all format in most cases in single-mode web, paper, and telephone surveys and in multiple or mixed-mode surveys. However, we can conceive of situations in which the forced-choice

format may not be ideal. For example, in a factual question with many response options the forced-choice format may be tedious. That said, in most instances the forced-choice format would seem to be more desirable than the check-all format, especially for opinion-based questions.

The larger implication of these results is that changes made to accommodate one particular survey mode may significantly change the stimulus respondents receive, making results across modes incomparable. Now that surveyors are increasingly relying on mixed-mode data collection strategies, it is important to experimentally test mode-specific changes (e.g., forced-choice and check-all; polar and fully labeled scales, etc.) to ensure that respondents are receiving the same stimulus. The encouraging news for survey designers is that in the case of multiple-answer questions, the forced-choice format can be used effectively within and across survey modes.

Supplementary Data

Supplementary data are available online at <http://pubopq.oxfordjournals.org/>.

References

- de Leeuw, Edith D. 1992. *Data Quality in Mail, Telephone, and Face to Face Surveys*. Amsterdam: TT Publications.
- de Leeuw, Edith D. 2005. "To Mix or Not to Mix: Data Collection Modes in Surveys." *Journal of Official Statistics* 21(2):233–55.
- Dillman, Don A., Roberta L. Sangster, John Tarnai, and Todd H. Rockwood. 1996. "Understanding Differences in People's Answers to Telephone and Mail Surveys." In *New Directions for Evaluation Series, 70 (Advances in Survey Research)*, eds. M. T. Braverman and J. K. Slater, pp. 45–62. San Francisco: Jossey-Bass.
- Jordan, Lawrence A., Alfred C. Marcus, and Leo G. Reeder. 1980. "Response Styles in Telephone and Household Interviewing: A Field Experiment." *Public Opinion Quarterly* 44(2):210–22.
- 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. 1999. "Survey Research." *Annual Review of Psychology* 50:537–67.
- 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.
- Nicolaas, Gerry, and Sarah Tipping. 2006. "Mode Effects in Social Capital Surveys." *Survey Methodology Bulletin*, pp. 56–74, Special ed. No. 58. England: Office for National Statistics.
- 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.
- Schwarz, Norbert, Fritz Strack, Hans-J. Hippler, and George Bishop. 1991. "The Impact of Administration Mode on Response Effects in Survey Measurement." *Applied Cognitive Psychology* 5:193–212.
- 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. Pullman, WA. www.sesrc.wsu.edu/dillman/papers.htm (retrieved May 6, 2007).
- Smyth, Jolene D., Don A. Dillman, Leah Melani Christian, and Michael J. Stern. 2006. “Comparing Check-All and Forced-Choice Question Formats in Web Surveys.” *Public Opinion Quarterly* 70:66–77.
- Sudman, Seymour, and Norman M. Bradburn. 1982. *Asking Questions*. San Francisco, CA: Jossey-Bass.
- Thomas, Randall K., and Jonathan D. Klein. 2006. “Merely Incidental?: Effects of Response Format on Self-reported Behavior.” *Journal of Official Statistics* 22:221–44.