

REDUCING MODE EFFECTS IN "MARK ALL THAT APPLY" QUESTIONS

Geraldine M. Mooney and Barbara Lepidus Carlson

Geraldine M. Mooney, Mathematica Policy Research, P.O. Box 2393, Princeton, New Jersey 08543-2393

Key Words: Primacy Effects, Recency Effects

Introduction

Using data from a mode effects study conducted as part of the 1993 National Survey of College Graduates, this paper looks at the extent to which mode effects in "Mark All That Apply" questions are minimized when such questions are reformatted to require a "yes or no" response after each response alternative.

A "Mark All That Apply" format, upon reading the question, asks respondents to look over a list of response categories and mark only those categories that apply. Because it is viewed as an effective format for minimizing respondent burden, it is frequently used. One serious shortcoming, however, is its vulnerability to mode effects. More specifically, visually presented "Mark All That Apply" questions in self-administered instruments are prone to "primacy effects"-- the tendency to select response alternatives from the top of a list, while orally presented (telephone-administered) "Mark All That Apply" questions are prone to "recency effects" or the tendency to mark items toward the bottom of the list.

Items placed early in a list have a definite advantage in self-administered questionnaires. As discussed by Tourangeau (1984), respondents work through four stages of cognitive processing when responding to a question: (1) comprehension-- interpreting the meaning of the question; (2) referral--retrieving relevant material from memory; (3) judgement--using the relevant material from memory to formulate a response; and (4) reporting--responding in a manner consistent with prior answers or with regard to other factors such as social desirability. To give every question in an entire interview or questionnaire such careful and deliberate consideration would require more time and effort than most respondents are willing or able to provide. Consequently, respondents generally seek to minimize their burden by "weak satisficing" or selecting the first reasonable response from the list of alternatives presented. (Schuman and Presser 1981, Krosnick and Alwin 1987, Krosnick 1991). This means initially listed response categories in a "Mark All That Apply" format are often more carefully considered than later ones, and thoughts relevant to those initial categories often influence how later response categories are perceived.

The prevalence of recency effects in telephone-administered questionnaires is influenced by the limitations of memory and the fact that respondents to telephone surveys tend to feel they have less control over the pace of the interview than do respondents to self-administered questionnaires. Since not prolonging the call seems to be an unspoken rule of telephone interviewing, telephone respondents are often less likely than respondents to a self-administered questionnaire to stop and ponder a particular response category or ask to have categories reread.

Similarly, with respect to memory, it appears most people cannot retain more than two or three response categories in memory at a time, again giving the last categories heard an advantage over response categories read earlier in a list. Cumulatively, both of these factors tend to favor making selections from among the final categories read (Schuman and Presser 1981, Tarnai and Dillman, Krosnick 1991, Krosnick and Alwin 1987, Swartz and Strack 1991).

Sudman and Bradburn (1982) recognized the shortcomings of "Mark All That Apply" questions more than a decade ago. They suggested data quality would improve if these questions were reformatted such that respondents were asked to indicate a negative or affirmative response to each individual response category, as opposed to only marking those that applied. While "more cumbersome," they argued it would encourage respondents to more thoroughly read and consider the response categories (Ibid, page 168).

Interestingly, very little research has been conducted on this issue. The only directly relevant research we uncovered was a recent experiment conducted by Rasinski, Mingay and Bradburn (Rasinski et al. 1994). As part of a self-administered questionnaire pretest, half the sample members were presented three questions formatted with "mark all that apply" instructions. The other half were presented the same three questions in a form that asked for a "yes or no" response after each item. Among the three questions, one had four response categories, one had twelve and one had twenty. They found the "yes/no" format produced more "yes" responses, but it was not clear whether this reflected better reporting accuracy or overreporting. The two items with fewest response categories did not show primacy effects in the "yes/no" format, but primacy was still evident in the one item with 20 response categories. The authors, however, felt these results should be taken with some caution given the limited number of items being tested.

Method

The Sample. This research took a slightly different tack. Using data from a mode effects study conducted as part of the 1993 National Survey of College Graduates (NSCG)¹, we compared the mail and telephone data from six questions purposively rewritten with the "yes/no" format in order to minimize mode effects. The NSCG is a mixed-mode survey--mail with telephone and in-person follow-up. As part of a larger evaluation of data quality, approximately 5,000 sample members were randomly drawn from the

¹The National Survey of College Graduates is sponsored by the National Science Foundation, with data collected by the U.S. Bureau of the Census. The sampling frame is all 1990 decennial census long form respondents who have at least a Bachelor's degree. The NSCG targets those college graduates who have been trained in, and/or are working in, a science or engineering field.

mail sample and initially contacted by telephone, rather than mail, to assess mode effects. For this paper, the “mail” group consists of those persons selected as part of the larger (mail) sample who responded by mail (i.e., those that required telephone or in-person follow-up were excluded). The computer-assisted telephone interviewing, or “CATI,” group consists of those persons selected for the mode effects study who responded either by CATI or by in-person interviewer follow-up.

The Questions. Ideally, we wanted questions positioned throughout the questionnaire and questions with a varying number of response categories. We considered position important because questions toward the start of the questionnaire will probably be read more carefully or, conversely, items toward the end will be more prone to satisficing. We also assumed the number of response categories would impact mode effects since fewer response categories should be less subject to memory limitations and cognitively less demanding than longer lists. Thus, we expect questions near the start of the questionnaire or those with fewer response alternatives to be less prone to primacy or recency.

As shown in Table 1, the six “yes/no” formatted items were well spaced throughout the questionnaire. Among the 12 pages of questions, two items were positioned near the beginning of the questionnaire, three were positioned in the middle and one came at the end of the questionnaire. The number of response categories also varied appropriately. The number ranged from four to 14, with most falling in between with either seven or nine response categories.

The Performance Measures. If the “yes/no” format is successful at reducing mode effects, the “yes” responses in these reformatted “Mark All That Apply” questions should be similarly distributed in both modes-- mail questionnaire responses should not be more clustered toward the top half of the response list than those of CATI responses.

To test this, we compared several performance measures for each question by mode:

- Mean number of “yes” responses, and the mean number of “yes” responses in the first and second halves of the response category lists (with seven categories or nine, we counted three and four in the top half, respectively)
- The proportion of “yes” responses in the first half of the response list
- Ranking response categories by number of yes responses
- The combined effect of position, number of response categories, and cognitive burden

To evaluate the joint effect of position, number of response categories, and cognitive burden, we developed an index that combined these three characteristics. This index was used to detect patterns in the distribution of yes responses between modes; that is, which questions were still most likely to suffer primacy or recency effects.

Results

Mean Number of “Yes” Responses. Table 2 shows the mean number of yes responses by question and mode for: (1) the first half of each response category list, (2) the second half of each response category list and (3) the entire list. As shown in Table 2, the tendency for mail “yes” responses to cluster at the top of a list and telephone “yes” responses to cluster toward the bottom of a list did not occur. In fact, regardless of a question’s position in the questionnaire or the number of response categories, administering the question by telephone always resulted in more yes responses.

The fact that administering these reformatted “Mark All That Apply” questions by telephone resulted in more “yes” responses is not too surprising. The “yes/no” format requests a response for each response category. Since telephone interviewers are required to read and record a response for each category, telephone respondents have little choice but to listen and respond. Mail respondents, on the other hand, can stop or succumb to weak satisficing at any point. Despite this option to stop among mail respondents, the difference in the mean number of yes responses by mode was always less than a single yes response. The magnitude of the differences varied little between the top half and the bottom half of each list-- between .01 to .44 of a yes response for the top half and .04 to .33 of a yes response difference in the second half.

For four of the six questions, both mail and CATI respondents averaged slightly more “yes” responses in the first half of each list. This may be due to the fact that the responses expected to be more commonly chosen were placed first in the list. Question C6 shows a hint of primacy and recency effects--the mail respondents had slightly more yes responses in the first half than the second half, and the CATI respondents had slightly more yes responses in the second half. Among mail respondents, only one question (D8) averaged more yes responses in the second half of the list than in the first half, while the CATI respondents averaged more yes responses in the first half of the list. This is surprising since D8 has nine response categories and is situated near the end of the questionnaire. With a “Mark All That Apply” format we would have ordinarily expected such a self-administered question to be particularly vulnerable to respondent fatigue and satisficing. The fact this did not happen suggests the yes/no format is causing more respondents to read through the entire list of response categories.

On the other hand, given these sample sizes (15,316 to 92,206 for the mail), even very small differences are almost always highly statistically significant (with p-values approaching zero), indicating that observed differences were almost certainly not due to sampling error. Only questions A20 and B10 had modal differences that were not statistically significant. Because A20 is positioned near the start of the questionnaire, when respondents may still be reading questions more carefully, this care could be causing A20 to be read in a manner that more closely replicates the telephone administration.

Proportion of Yes Responses in the First Half of the List. As noted earlier, reformatting the “Mark All That

Apply” questions into the yes/no format appears to have minimized the expected mode effect but another mode effect emerged--- administering these questions by telephone produced an overall increase in the average number of yes responses. To control for this phenomenon, we also examined whether primacy and recency effects were apparent if we controlled for the number of yes responses by looking at the proportion of yes responses marked in the first half of each response category list in each of the two modes.

As shown in Table 3, CATI respondents marked proportionally more yes responses in the first half than did mail respondents (one to two percentage points) for four of the six questions. In three of those instances, however, the differences were not statistically significant. Although self-administration had produced proportionally more yes responses in the first half compared with CATI respondents in only two of the six questions, these questions indicated relatively larger differences--three to five percentage points. Again, however, neither position in the questionnaire or number of response categories indicated where these greater differences would lie. For example, the two questions with the greatest differences across mode were positioned in the middle of the questionnaire and one item (C5) has only four response categories. Conversely, A20 (seven response categories), positioned at the start of the questionnaire and D8 (nine response categories), positioned near the end, had differences that were not statistically significant.

Ranking Response Categories by the Percent of “Yes” Responses. If the yes/no format is successful at minimizing mode effects, we should see very similar rankings across mode of the prevalence with which responses are marked. As shown in Table 4, the rankings differed little across mode. For two of the six questions (C6 and D8), the rankings were identical. For three of the remaining four questions (A20, B10, C5), the rankings were off for two of the response categories. Question A22, with its 14 response categories, indicated the largest number of differences, with three of the response categories having different rankings.

An Index of Combined Effects. Given that neither position or number of response categories appeared definitive for identifying those questions most likely to be affected by mode effects, we attempted to measure the combined effect of three factors: (1) position in the questionnaire, (2) number of response categories, and (3) cognitive burden. Table 1 shows the position of each question in the questionnaire and the number of response categories. Table 5 shows how we scored each question on cognitive burden.

To score position in questionnaire, we divided the twelve pages of questions into four equal sections, with each section worth one point. In addition, an extra half point of burden was added if the question was the second yes/no formatted question on a page. The scale had a range of 1 to 4.5 (see Table 6).

We scored number of response categories by aggregating response categories in units of five. The first group, with one to five response categories, was given a score of “1”

and 1.5 points was added for each additional cluster of five response categories. For these questions, the individual question scores ranged from 1 to 4 (see Table 6).

We defined cognitive burden as having three primary dimensions: length of the recall period, degree of salience, and degree of judgement required. The three dimensions were scored individually on three-point scales using the following parameters:

- **Length of Recall Period:** Events occurring at the time the interview or within the past year were scored a “1” while events that could have happened as long as 20 years ago were scored a “3”
- **Salience:** Very salient items were scored a “1” (e.g., reasons for a fairly recent job change) while items with little salience were scored a “3”
- **Judgement:** Questions with well-defined response categories that asked for unambiguous information were scored a “1” while more ambiguous questions or response categories that required more thought (e.g., asking respondents to classify their reference week job according to a detailed list of prescribed work activities) were scored a “3”

The three scores were summed to create an overall cognitive burden score which could range from a low of 3 to a high of 9. As shown in Table 5, the cognitive burden scores clustered closely together toward the lower end of the scale. The questions varied the least with respect to salience and about equally for length of the recall period and degree of judgement.

Table 6 presents the individual scores for position, number of response categories, and cognitive burden and the combined index:

- **Ordering by position in the questionnaire** indicated that those items with the largest differences across mode tended to be positioned later in the questionnaire. The one notable exception was A22, a question found on page 3, but one with 14 response categories. A22 had the second largest difference by mode, second only to D8, the question placed last in the questionnaire among the yes/no reformatted questions. Clearly position is important, but its effects can be mitigated by a particularly burdensome number of response categories.
- **Ordering by number of response categories** produced more mixed results. Although two questions with the largest number of categories (A22 and D8) had two of the largest differences by mode, B10, the other question with nine response categories, had the smallest across mode difference.
- **Ordering by cognitive burden** also produced more mixed results. Although our cognitive burden measure assigned the lowest burden score to the question with the smallest modal difference in the mean number of yes responses, it also judged A20 and D8 to be of equal

burden (for different reasons) even though D8 had the greatest difference in the mean number of yeses by mode and A20 had the second smallest difference.

Since we assumed the three factors to be of equal weight, we simply added the three scores to create the combined index. A question with the lowest score is considered least likely to suffer from primacy or recency, while a question with the highest score is considered the most vulnerable. Combining these three factors into the index improved their ability to differentiate differences in the mean number of yeses by mode. Except for the fact that the two questions with the smallest differences were reversed, increases in the combined index score paralleled increases in the difference of average number of yes responses by mode.

Discussion

“Mark All That Apply” questions are particularly susceptible to the mode effects of primacy when self-administered, and recency when administered by telephone.

This study looked at whether changing the “Mark All That Apply” instruction into a format that requires a “yes” or “no” after each response category mitigates these mode effects. Since we could not administer each question by mail and telephone in both formats, we attempted to view the potential impact of the reformatting indirectly by looking for the absence of primacy and recency when the reformatted questions were administered by mail and by telephone.

For primacy we examined the mean number of yes responses in the first half of the response category list by mode and, for recency, we examined the mean number of yes responses in the second half of the list by mode. Although primacy among mail respondents and recency among CATI respondents was not evident, we found another mode effect -- the fact that administering these questions by CATI produced more “yes” responses in the first half of the list and more yeses in the second half as well. While this trend was apparent in all six questions, differences by mode were generally small, but statistically significant due to the very large sample sizes.

This effect is likely the result of the more structured environment of an interviewer-administered questionnaire.

Because the CATI interviewers are required to read and record a response for each response category, CATI respondents are led by the interviewer through each response category. Mail respondents, on the other hand, do not have this interviewer intervention. Although instructed to mark a yes or no for each response category, they can stop marking responses at any point. However, as noted by Rasinski et al. (1994), without outside verification, it is difficult to ascertain if an increase in the number of yes responses reflects greater accuracy or overreporting. Although CATI produced more yeses, the fact that primacy was not evident among mail respondents suggests the yes/no format was causing mail respondents to read the list of response alternatives more completely. This is reinforced by the fact that one yes/no question with nine response categories averaged more yes responses in the

second half of the list than in the first half for mail respondents, while the opposite was true for CATI respondents, even though this question was positioned near the end of the questionnaire.

Using the proportion of yes responses marked in the first half of the response category list, a measure less sensitive to the overall number of yes responses marked, we found that CATI respondents had proportionally more yes responses in the top half of the list than did mail respondents for four of the six questions, but these differences were not statistically significant for three of the four items. Although self-administration produced a small but statistically significant increase in the proportion of yes responses in the first half over the proportion found in CATI administration in the remaining two questions, suggesting a subtle primacy effect among mail respondents, it appears the overall analytical implications of these differences in the mean number or proportion of yes responses by mode are minor. In support of this, we also compared how response categories ranked by the number of yes responses, across mode. The yes/no format produced identical or very similar rankings for five of the six questions.

Finally, we looked at the impact of three question characteristics (position in the questionnaire, number of response categories, and cognitive burden) known to affect the likelihood of primacy and recency. In general, we found neither position in the questionnaire nor number of response categories fully explained modal differences in the total number or proportion of yes responses -- suggesting that the reformatting might have mitigated these particular effects. However, cognitive burden and using an index that combined the impact of these three question characteristics did seem to be positively correlated with larger differences by mode, suggesting that the reformatting had less of an impact on questions that were more difficult to answer.

Limitations and Directions for Future Research

When interpreting the results of this paper, it should be noted that it was conducted on one survey, on only six questions, and on a relatively sophisticated population (college graduates in the fields of science or engineering). Similar evaluations should be done using other surveys with a variety of populations. Ideally, a direct comparison of “Mark All That Apply” and “Yes/No” formats would be carried out as well.

Because we did not have the data in hand at the time this research was conducted, our ability to carry out more sophisticated analysis was limited. We plan to obtain the data and carry out factor analyses to determine the interactions among the question characteristics (position, number of responses, and cognitive burden). Another direction for future research is an attempt to determine whether the higher number of yes responses found for interviewer-administered surveys reflects more accurate reporting or overreporting.

Appendix

The six questions examined were:

A20: “Did these factors influence your decision to work in an area outside of your highest degree field?” had six categories, plus an “Other - Specify” response;

- A22 "Did the following work activities occupy 10 percent or more of your time during a typical work week on this [principal] job?" had thirteen categories, plus "Other - Specify"
- B10 "Did these factors influence your decision to change employers or occupations between April 1988 and the week of April 15, 1993?" had eight categories, plus "Other - Specify"
- C5 "During the past year, in which of the following areas did you attend work-related workshops, seminars, or other work-related training activities?" had three categories, plus "Other - Specify";
- C6 "For which of the following reasons did you attend training activities during the past year?" had six categories, plus "Other - Specify"
- D8 "For which of the following reasons did you take college courses between completing your most recent degree and the week of April 15, 1993" had eight categories, plus "Other - Specify."

Krosnick JA (1991). "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology*, Vol. 5, 213-236.

Rasinski KA, D Mingay, and NM Bradburn (1994). "Do Respondents Really 'Mark All That Apply' on Self-Administered Questions?" *Public Opinion Quarterly*, Vol. 58: 400-408.

Schuman and Presser (1981). "Questions and Answers in Attitude Surveys." Academic Press: Orlando, FL.

Sudman and Bradburn (1982). "Asking Questions." Jossey-Bass, Inc: San Francisco.

Sudman S, NM Bradburn, R Schwarz (eds). *Thinking About Answers*. Jossey-Bass, Inc.: San Francisco.

Swartz and Strack (1991). "The Impact of Administration Mode on Response Effects in Survey Measurement." *Applied Cognitive Psychology*, Vol. 5, 193-212.

Tarnai and Dillman (1992). "Questionnaire Context as a Source of Response Differences in Mail and Telephone Survey." In Norbert Schwarz and Seymour Sudman (eds): *Context Effects in Social and Psychological Research*. New York: Springer-Verlag (pp. 115-124).

Tourangeau R (1984). "Cognitive Sciences and Survey Methods." In T. Jabine, M. Straf, J. Tanur, and R. Tourangeau (eds), *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines* (pp. 73-100). Washington, DC: National Academy Press.

References

Krosnick JA and DF Alwin (1987). "An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement." *Public Opinion Quarterly*, Vol. 51: 201-219.

TABLE 1
QUESTIONNAIRE ITEMS EXAMINED

Question Number	Position in Questionnaire	Number of Categories (Including "Other")	Number of CATI Respondents	Number of Mail Respondents
A20	Page 3 of 12	7	584	15,316
A22	Page 3 of 12	14	3,193	92,206
B10	Page 7 of 12	9	1,344	32,855
C5	Page 8 of 12	4	2,154	69,005
C6	Page 8 of 12	7	2,154	69,016
D8	Page 10 of 12	9	1,472	50,381

TABLE 2
MEAN NUMBER OF "YES" RESPONSES
ORDERED BY POSITION IN QUESTIONNAIRE

Question Number		Mail		CATI		Difference (Mail-CATI)		Signif of Of 2-Sided Test $\alpha = .10$
		Mean	s.e.	Mean	s.e.	Mean	s.e.	
A20	Yeses in 1st half	1.379	(.008)	1.404	(.042)	-0.025	(.043)	NS
	Yeses in 2nd half	1.215	(.006)	1.271	(.035)	-0.056	(.035)	NS
	Total yeses	2.593	(.011)	2.675	(.055)	-0.082	(.056)	NS
A22	Yeses in 1st half	1.718	(.005)	2.044	(.028)	-0.326	(.029)	S
	Yeses in 2nd half	1.710	(.004)	1.925	(.022)	-0.215	(.023)	S
	Total yeses	3.427	(.007)	3.970	(.040)	-0.543	(.041)	S
B10	Yeses in 1st half	1.576	(.007)	1.584	(.031)	-0.008	(.032)	NS
	Yeses in 2nd half	0.679	(.004)	0.721	(.018)	-0.042	(.019)	S
	Total yeses	2.254	(.007)	2.305	(.034)	-0.051	(.035)	NS
C5	Yeses in 1st half	1.029	(.002)	1.090	(.013)	-0.061	(.013)	S
	Yeses in 2nd half	0.496	(.002)	0.688	(.014)	-0.192	(.014)	S
	Total yeses	1.525	(.003)	1.777	(.018)	-0.252	(.019)	S
C6	Yeses in 1st half	1.274	(.002)	1.353	(.013)	-0.079	(.014)	S
	Yeses in 2nd half	1.115	(.004)	1.435	(.021)	-0.320	(.022)	S
	Total yeses	2.389	(.005)	2.787	(.027)	-0.398	(.028)	S
D8	Yeses in 1st half	1.395	(.004)	1.838	(.026)	-0.443	(.027)	S
	Yeses in 2nd half	1.421	(.005)	1.754	(.026)	-0.333	(.027)	S
	Total yeses	2.816	(.007)	3.592	(.041)	-0.776	(.042)	S

TABLE 3
PROPORTION OF YES RESPONSES IN FIRST HALF OF RESPONSE CATEGORY LIST

Question	Mail	CATI	Difference Mail-CATI	s.e.	Significance $\alpha = .10$
A22	.434	.458	-.024	.0055	S
B10	.619	.633	-.014	.0105	NS
A20	.474	.484	-.010	.0136	NS
D8	.501	.509	-.008	.0061	NS
C5	.682	.650	.032	.0076	S
C6	.595	.542	.053	.0057	S

TABLE 4 is available from the authors upon request.

TABLE 5
SCORING COGNITIVE BURDEN¹

Question Number	Scores			TOTAL
	Length of Recall	Degree of Salience	Degree of Judgment	
A20	3	1	1	5
A22	1	2	3	6
B10	2	1	1	4
C5	1	2	3	6
C6	1	2	2	5
D8	3	1	1	5

¹A score of one indicates the least amount of cognitive burden.

TABLE 6
DIFFERENCES IN THE MEAN NUMBER OF YESES ORDERED BY POSITION,
NUMBER OF RESPONSE CATEGORIES AND COGNITIVE BURDEN

A. Characteristic: Position in the Questionnaire			
Question	Score	Difference	
A20	1	-.082	
A22	1.5	-.543 *	
B10	3	-.051	
C5	3	-.252 *	
C6	3.5	-.398 *	
D8	4	-.776 *	
B. Characteristic: Number of Response Categories			
Question	Score	Difference	
C5	1	-.252 *	
A20	2.5	-.082	
C6	2.5	-.398 *	
B10	2.5	-.051	
D8	2.5	-.776 *	
A22	4	-.543 *	
C. Characteristic: Cognitive Burden			
Question	Score	Difference	
B10	4	-.051	
A20	5	-.082	
C6	5	-.398 *	
D8	5	-.776 *	
C5	6	-.252 *	
A22	6	-.543 *	
D. Index of Position, Number of Response Categories and Cognitive Burden			
Question	Score	Difference	
A20	8.5	-.082	
B10	9.5	-.051	
C5	10	-.252*	
C6	11	-.398*	
A22	11.5	-.543*	
D8	11.5	-.776*	

*Significant at $\alpha = .10$