Evaluation and Program Planning, Vol. 13, pp. 365-371, 1990
Printed in the USA.

0149-7189/90 $3.00 + .00

CAN RESPONSE ORDER BIAS EVALUATIONS?

GLENN D. ISRAEL and C. L. TAYLOR

University of Florida

ABSTRACT

Response-order effects can bias evaluations. Measurement error is introduced by order effects, and this, in turn, can affect associations between variables. This study attempts to identify types of questions in mail questionnaires which are susceptible to order effects and illustrates the consequences of response-order effects on associations. Data from a Cooperative Extension Service evaluation showed order effects for multiple response and attributive questions but not for single response ones. Aspects of question complexity and social desirability may contribute to order effects. In at least one instance, response-order effects were found to influence associations between items and thus, the substantive conclusions.

CAN RESPONSE-ORDER BIAS EVALUATIONS?

Although much research has been directed at assessing the effects of the order and wording of questions on responses in evaluation surveys, less attention has been directed toward bias which can occur from the order of response alternatives in questions. Response-order effects in closed survey questions are said to occur when the order in which the alternatives are listed influences respondents' choices (Schuman & Presser, 1981). The consequences of response-order effects are two-fold. First, the distribution of responses can be affected, thereby introducing measurement error in variables for evaluation studies. For panel studies, comparability requires the same response sequences be used (Carp, 1974; see also Kraut, Wolfson, & Rothenberg, 1975), but the possibility of measurement error for each instance of data collection remains. Second, response-order effects can influence associations between variables (Kalton, Collins, & Brook, 1978). Measurement error that is introduced by response-order effects can attenuate relationships, resulting in reduced significance and/or the nonrejection of null hypotheses when they should be rejected. As Kalton et al. (1978) note, potential bias due to response order deserves wider attention by evaluators than has previously been the case.

To date, most studies of response-order effects used telephone and face-to-face interviews and little attention has been directed at mail questionnaires (see, for example, Bishop, Hippler, Schwarz, & Strack, 1988; Carp, 1974; Kraut, Wolfson, & Rothenberg, 1975; Krosnick & Alwin, 1987; Schuman & Presser, 1981). However, mail surveys continue to be an important means for gathering evaluation data. The purpose of this paper is to examine the occurrence of response-order effects in a mail survey and to identify which types of questions are susceptible to these effects. The potential for response order to influence substantive conclusions about the effectiveness of a program is illustrated.

ISSUES FOR STUDY

Among the most commonly mentioned response-order effects are primacy and recency effects. A primacy effect is the tendency to select a response near the beginning of a list of alternatives. Likewise, a recency effect is the tendency to select the responses near the end of a list of alternatives. A number of studies using long lists of response choices have found a primacy effect (Becker, 1954; Brook & Upton, 1974; Krosnick & Al-

Revision of the paper presented at the annual meeting of the American Evaluation Association in Boston, MA, October 1987. The authors would like to thank the anonymous reviewers for comments on an earlier draft of the paper.

Requests for reprints should be sent to Glenn D. Israel, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL 32601.

win, 1987; Mueller, 1970). In a study of radio show preferences, Becker (1954) found that the chances of being selected as one of the five most preferred (of 16 programs on the cardlist) increased as the program type moved from the sixteenth position to the first position on a printed checklist. The voting studies by Brook and Upton (1974) and Mueller (1970) both found that in nonpartisan elections with numerous nominees, candidates whose names appeared near the top of the ballot gained votes as a consequence of that position.

Although primacy effects have tended to be found for printed lists, especially those with numerous response options, recency effects have tended to be found for orally administered questions (Krosnick & Alwin, 1987; Schuman & Presser, 1981). Primacy effects in written lists are said to occur when respondents select the first satisfactory response or responses from the list of alternatives, rather than selecting the optimal response (Krosnick & Alwin, 1987). Respondents who are low in cognitive sophistication are more likely than those high in cognitive sophistication to use the satisfying strategy in selecting responses. Recency effects tend to occur from limitations in short-term memory to recall earlier alternatives when no written reminders are present.

Response-order effects resulting from different response strategies may be minimized in self-administered surveys because respondents are likely to have more time to think about answering each question (Bishop et al., 1988). An optimizing strategy might be used by more respondents to a mail questionnaire and this can yield more reliable data than would be obtained by other data collection methods (e.g., telephone survey or personal interview).

Although self-administered surveys may minimize response-order effects, those that remain can result from a number of factors. The complexity of the question responses and the extent to which opinions are defined have been suggested as reasons for response-order effects (Becker, 1954; Rugg & Cantril, 1944; cf. Schuman & Presser, 1981). There are many aspects to complexity, including the number of alternatives, physical presentation (e.g., one column or multiple columns), and the number of responses to be selected (e.g., single or multiple). Each of these can have an effect on the use of a satisfying or optimizing mode to respond to questions.

Order effects also can occur from social desirability, the tendency of some respondents to be chronic "yea-sayers" or "nay-sayers" on Likert-type items, and the nature of the question (see Carp, 1974). A social desirability effect can occur in questions asking about attribution of program effects by listing the sponsoring agency in a prominent position (i.e., first or last). For example, Cooperative Extension Service surveys which ask where the respondent obtained educational information might induce a larger number of respondents to select "Extension" from the list of alternatives if that response were listed first, simply because that organization sent the survey. Previous contact between respondents and Extension personnel can foster the development of an exchange relationship (Homans, 1961) in which support for the organization is expected from its clientele in return for the provision of educational information.

The nature of the question also can influence whether or not response-order effects might occur. In a survey of elderly persons on transportation issues, Carp (1974) found response-order effects occurred for evaluative questions consistently, but infrequently for factual questions. An extension of this finding would suggest that questions which include response alternatives that are ambiguous can produce "wording effects" (cf. Bishop et al., 1988) which influence the appearance of response-order effects.

This study examines the influence of response order on single response questions, multiple response questions, and attributive questions to see which types are likely to show order differences. The effect of the number of alternatives listed (from three categories for several questions to nine for another) and whether single or multiple categories are selected for a single question is examined to assess the effects of question complexity. In addition, questions about the attribution of services form a basis to examine response-order bias from social desirability.

METHODS

The source of data for this study is a survey conducted by the Florida Cooperative Extension Service in 1986 to evaluate the impact of educational programs concerning the production and management of beef cattle. County Extension Agents obtained the data from producers using mail questionnaires. From their county mailing lists, agents identified 2458 beef producers in nine counties in the Greater Tampa Bay Area, Florida. A random sample of 344 was selected. Using mailing techniques from the "Total Design Method" (Dillman, 1978), 168 usable questionnaires were obtained. This represents at 48.8% response rate. The resulting error rate for a sample of this size is ±7.3% within a 95% confidence interval.

Two survey forms were utilized for response-order experiments. The forms were identical in format and question order. The only differences were the order of response options for 10 questions on the instrument. At

least one question was located on each page of the four-page questionnaire. The forms were randomly assigned to potential respondents by flipping a coin. A total of 76 respondents returned form "A" and 92 respondents returned form "B."[1] One of the 10 questions was eliminated from the study because too many respondents were screened out by the immediately preceding question. The specific questions for the remaining nine items are presented in the accompanying tables. Five of the questions solicit a single answer and four ask for multiple responses, with one of each type also being an ascriptive question.

The analysis compares the distributions of responses from the two forms using the chi-square statistic. A single chi-square test was applied for questions with mutually exclusive response categories. For questions in which multiple responses could be selected, chi-square was calculated for each category to test the hypothesis of equal proportions.

RESULTS

Of the nine questions about beef management and production practices tested, two suggest that the order of the alternatives influences the distribution of responses. The results for the single response questions are presented first, followed by the multiple response questions, and then the attributive questions.

Single Response Questions

Producers were asked about their plans for their beef operation over the next 5 years. As shown in Table 1, no significant difference is apparent between the two listings of response alternatives. For each of the four alternatives, differences between the first order of listing and the second are 1.6 percentage points or less.

The results from Table 2 for the question in which producers evaluate the importance of making a profit are similar to the first. Producers were asked how important was it for their beef operation to make a profit. Nonsignificant order effects are apparent. Percentage differences are 2.5 or less.

A third single response question was employed when producers were asked about the adequacy of their winter forage program. As shown in Table 3, no significant difference was apparent when the order of the three alternatives was changed, although the percentage who responded "yes" increased 10.8 points when the order of the response was changed from first to second in the list.

The last single response question included in the Beef-Forage survey asked producers to identify the single most important problem facing the beef cattle industry. The results are shown in Table 4. The difference between the distribution of the two listings of responses is not significant. The largest percentage difference is 6.5 points for the "diet/health concern" option, followed by 5.3 percentage points for "price rancher receives." Neither this question nor the other three single response questions were found to have response-order effects.

Multiple Response Questions

Beef producers also were asked for information about the methods that they used for parasite control (see Table 5). This question was more complex for two reasons. First, the item involving the alternative response

TABLE 1
RESPONSE-ORDER EFFECTS OF FUTURE PLANS OF BEEF PRODUCERS

The question read: Over the next 5 years, are there plans to:

Alternatives	Order of Alternatives ABCD	DCBA
A. Increase the size of this beef operation	25.3%	25.3%
B. Reduce the size of this beef operation	5.3	6.6
C. Maintain about the same size beef operation	60.0	60.4
D. Get out of the beef business	9.3	7.7
	99.9	100.0
	(75)	(91)

chi-square = 0.241, $df = 3$, n.s.

TABLE 2
RESPONSE-ORDER EFFECTS ON IMPORTANCE THAT THE BEEF OPERATION MAKES A PROFIT

The question read: How important is it for this beef operation to make a profit?

Alternatives	Order of Alternatives ABC	CBA
A. Very important	64.5%	62.6%
B. Somewhat important	31.6	34.1
C. Not important	4.0	3.3
	100.1	100.0
	(76)	(91)

chi-square = 0.149, $df = 2$, n.s.

[1]Although the sample size for the two forms differed, the response rate was nearly identical (47.5% for form A and 50.0% for form B). Respondents were found to differ on the number of years in the cattle business, but not on the size of the operation (in acres), herd size, farm structure, off-farm employment, and whether the farm is a major source of income. Logistic regression analysis (not shown here) indicates that differences in longevity in the cattle business between producers receiving form A and those with form B had no statistically significant effect on the findings reported later in the paper.

TABLE 3
RESPONSE-ORDER EFFECTS ON ADEQUACY OF WINTER FEEDING PROGRAMS OF BEEF PRODUCERS

The question read: Do you feel your winter feeding program is adequate for greatest net returns?

Alternatives	Order of Alternative ABC	CAB
A. Yes	39.2%	50.0%
B. No	27.0	22.6
C. Uncertain	33.8	27.4
	100.1	100.0
	(74)	(84)

chi-square = 1.864, *df* = 2, n.s.

TABLE 4
RESPONSE-ORDER EFFECTS ON PERCEIVED PROBLEMS FACING THE BEEF CATTLE INDUSTRY

The question read: What do you feel is the single most important problem facing the beef cattle industry?

Alternatives	Order of Alternative ABCDEFGHI	GHEDFCBAI
A. Retail price of beef	6.0%	1.3%
B. Price rancher receives	47.8	42.5
C. Demand for beef	25.4	22.5
D. Environmental issues	0.0	1.3
E. Animal welfare issues	0.0	0.0
F. Governmental regulation	10.5	10.0
G. Diet/health concerns	6.0	12.5
H. Zoning issues	1.5	0.0
I. Other	3.0	10.0
	100.1	100.1
	(67)	(80)

chi-square = 9.048, *df* = 7, n.s.

listings followed a screening question. Second, the alternatives were listed in two columns, with four responses in the first column and three in the second. The order of listing was changed within the columns only and not between the columns. Despite the complexity of the presentation of the responses, no significant differences were found. The largest difference was 7.0 percentage points.

As a follow-up question to how beef producers felt about the adequacy of their forage program (shown earlier in Table 3), they were asked to identify all the types of forages that were used on their farm. Respondents were instructed to check all that applied. Changing the order in which the alternatives were presented resulted in significant differences ($p < .05$) for one of the six responses (see Table 6). When "native range" was moved from the first poisition to the fourth, the percentage of respondents who selected that item decreased from 69.7% to 30.4%. There appears to be a primacy effect associated with this response alternative. Although "deferred grazing" remained in the same position on both forms (second in the list of alternatives), the percentage of producers who selected that item increased from 36.8% to 47.8%. This increase is apparently a result of replacing a high response item "native range" with a low response item "silage" in the immediately prior position. Differences among the remaining items are not significant, although the percentage selecting "winter pasture" increased 6.9 percentage points when its position was moved from fifth to third. In the second listing, "winter pasture" was listed prior to "native range" and apparently benefited from this relationship in the same way that "deferred grazing" did. Of note is the stability in the percentage selecting "hay," with 84.2% for the third position (form A) and 79.4%

TABLE 5
RESPONSE-ORDER EFFECTS ON PARASITE CONTROL PRACTICES USED BY BEEF PRODUCERS

Respondents who control for external parasites were asked: If yes, which of the following do you use? (check all that apply)

Alternatives	Order of Alternatives[a] ABCDEFG	CBADGFE	chi-square	*df*	prob.
A. Ear tags	25.0%[b]	18.0%	1.210	1	.271
B. Pour on	30.3	36.0	0.598	1	.439
C. Medicated feed	7.9	7.9	0.000	1	.994
D. Other	11.8	9.0	0.361	1	.548
E. Back rubber	9.2	13.5	0.734	1	.391
F. Dust bag	25.0	31.5	0.840	1	.359
G. Spray	55.3	61.8	0.723	1	.395
	(75)	(89)			

[a]Responses E, F, and G were listed in the second of two columns.
[b]Percentage selecting that response.

TABLE 6
RESPONSE-ORDER EFFECTS ON RESPONSES ABOUT WINTER FORAGE PRACTICES BY BEEF PRODUCERS

The question read: Which of the following forages are used during the winter months? (check all that apply)

Alternatives	Order of Alternatives ABCDEF	DBEACF	chi-square	*df*	prob.
A. Native range	69.7%[a]	30.4%	25.748	1	.000
B. Deferred grazing (save pasture for fall and winter use)	36.8	47.8	2.050	1	.152
C. Hay	84.2	79.4	0.654	1	.419
D. Silage	1.3	2.2	0.175	1	.676
E. Winter pasture (ryegrass, rye, wheat, oats)	29.0	35.9	0.906	1	.341
F. Other	14.5	15.2	0.018	1	.893
	(76)	(92)			

[a]Percentage selecting that response.

TABLE 7
RESPONSE-ORDER EFFECTS ON RESPONSES ABOUT USE OF PROTEIN SUPPLEMENTS BY BEEF PRODUCERS

The question read: Which of the following protein supplements are used during the winter months?

Alternatives	Order of Alternatives ABCDEF	FBCADE	chi-square	*df*	prob.
A. Molasses supplement	55.3%[a]	59.8%	0.348	1	.555
B. Range cubes	21.1	18.5	0.175	1	.676
C. Protein blocks	32.9	27.2	0.652	1	.420
D. Mineral containing protein	36.8	48.9	2.468	1	.116
E. Other	4.0	5.4	0.203	1	.652
F. None	5.3	1.1	2.514	1	.113
	(76)	(92)			

[a]Percentage selecting that response.

TABLE 8
RESPONSE-ORDER EFFECTS ON SOIL-TESTING ITEM

Respondents who use soil tests to guide their fertilizer and liming decisions were asked: Who tests your soil?

Alternatives	Order of Alternatives ABC	CBA
A. University of Florida	18.0%	15.0%
B. Fertilizer or lime company	74.0	73.3
C. Private laboratory	8.0	11.7
	100.0	100.0
	(50)	(60)

chi-square = 0.518, *df* = 2, n.s.

TABLE 9
RESPONSE-ORDER EFFECTS ON IDENTIFICATION OF SOURCE OF INFORMATION

The question read: Where, or to whom, do you look for information on beef production and/or management problems? (check all appropriate)

Alternatives	Order of Alternatives ABCDEFGH	CGEFBDAH	chi-square	*df*	prob.
A. Other cattlemen	69.3%[a]	62.0%	0.992	1	.319
B. Fertilizer, feed, machinery dealers, vaccine salesman	24.0	29.4	0.600	1	.438
C. Extension agent/ specialist	41.3	58.7	4.984	1	.026
D. Close relative (father, brother, etc.)	16.0	9.8	1.453	1	.228
E. Magazines	53.3	53.3	0.000	1	.993
F. Farm Organizations	16.0	20.7	0.592	1	.442
G. Veterinarian	52.0	44.6	0.915	1	.339
H. Other	5.3	4.4	0.088	1	.767
	(75)	(92)			

[a]Percentage selecting that response.

for the fifth position (form B). Clearly, order has an effect on the overall pattern of responses for this question.

The distribution of responses to another multiple response question is shown in Table 7. Beef producers were asked to identify which protein supplements to their forage were used during the winter months. No significant differences were found in the percentage selecting individual items, although there is a difference of 12.1 percentage points for one of the six alternatives. There does not appear to be a primacy effect on responses to this question, in contrast to that found for winter forage in Table 6.

Attributive Questions

The first of two attribution questions asked who performed soil tests for beef producers. The University of Florida is listed among the three alternatives (see Table 8). Since the Florida Cooperative Extension Service is housed at the University of Florida, a social acceptability effect might be anticipated. However, only a 3.0% advantage for the first position relative to the third is found for the "University of Florida" response. The presence of response-order bias due to social desirability is not supported by the responses to this question.

The second of the attribution questions asked respondents where or to whom they looked for information on beef management and/or production problems. In this question, the response option "Extension agent/specialist" is a potential source of social acceptability effects and, thus, response-order bias. As shown in Table 9, there is a significant difference. The item of central interest, "Extension agent/specialist," shows a significance ($p = .026$) primacy effect in the predicted direction. When the position of this option is moved from third to first, the percentage of beef producers who select that response increases from 41.3% to 58.7%. This suggests that a conservative approach be used in positioning response options for questions in which social desirability effects might occur.

The Influence of Response-Order Effects on Associations

In the previous section, order effects on the distribution of responses for individual questions have been examined. But as issue of greater importance is that of how response-order effects for one variable can result in different relationships with other variables, depending on the order in which response options are listed. The potential to reach substantively different conclusions solely as a result of response order is at issue here.

Given the above analysis, the relationship between using information[2] in selecting replacement herd bulls

[2]Information includes weaning weight, blood lines, and other relevant data.

and looking to Extension agents/specialists for information on beef management and/or production problems is used to illustrate the potential for response order to confound relationships of the type that are frequently examined in evaluation studies.

As shown in Table 10, the order in which the various sources of information are listed appears to create differences in the relationship between using information in selecting bulls and looking to Extension for information. When Extension is listed third among sources of information, there is essentially no relationship between looking to Extension for information and the use of information in selecting replacement bulls. In contrast, looking to Extension for information shows a strong positive effect on using information to select bulls when Extension is listed first among the alternatives.[3] The data indicates that Extension might be perceived to be effective in getting farmers to use information in selecting bulls on the basis of being listed as the first response alternative but might be viewed as ineffective when listed third among sources of information. Clearly, there is an order effect that is biasing this relationship.

TABLE 10
RESPONSE-ORDER EFFECTS ON THE RELATIONSHIP BETWEEN USE OF INFORMATION IN SELECTING BULLS AND LOOKING TO EXTENSION FOR INFORMATION

Use Information in Selecting Bulls	Response Order					
	First Position			Third Position		
	Look to Extension			Look to Extension		
	No	Yes	Total	No	Yes	Total
No	85.3%	39.6%	57.5%	74.4%	70.0%	72.6%
Yes	14.7	60.4	42.5	25.6	30.0	27.4
	100.0 (34)	100.0 (53)	100.0 (87)	100.0 (43)	100.0 (30)	100.0 (73)

CONCLUSIONS AND DISCUSSION

The findings suggest that response-order effects can be an important source of bias in evaluation studies using mail questionnaires. The order of response alternatives was found to affect both the distribution of responses for some items and the associations between these items and others. Reasons for response-order effects were supported by the findings in some cases but not in others. The idea of social desirability appears to warrant further investigation with one of the two questions of this type showing an order effect. In the question asking respondents about sources of information for management and production problems, Extension agents were identified by more respondents when listed in the first position, rather than the third position. One reason for this difference is that the more prominent response position may have attracted the attention of respondents to identify their relationship with the sponsoring agency, and thereby allowing respondents to meet expectations underlying the exchange relationship (Homans, 1961). Effects of this type are especially relevant to evaluators who are concerned about conducting studies which attribute impacts to specific organizations.

Evidence about the effect of complexity on response order also is mixed. Although all the questions were relatively short, no order effects were evident for the question which involved two columns of alternatives. Likewise, the number of alternatives in the list shows no relationship with the presence or absence of order effects. On the other hand, while items in which respondents were to select a single category showed no order effects, items which allowed multiple categories to be selected showed response-order effects in two of four experiments.

One question which demonstrated a clear order effect (forages used during the winter months, Table 6) suggests that ambiguity in the meaning of the response options to this question may be a cause of bias. If the term "native range" can be applied to a variety of pasture types by respondents, then the order in which this option is presented might influence the probability of selection. The order effect due to the wording of response categories can be as serious as that from the wording of questions (cf. Bishop et al., 1988). Thus, response alternatives need to be clear and matched with terms familiar to potential respondents.

The examination of order effects on associations serves to illustrate the practical application of response-order effects and the need for evaluators to concern themselves with this issue. It would seem that we are in danger of incorrectly estimating the impacts of programs by ignoring potential order effects while conducting evaluation studies. As Schuman and Presser (1981, p. 77) note, "greater understanding of these fun-

[3]Log-linear analysis reveals the presence of a significant three-way relationship between response-order position (P), looking to Extension for information (E), and using information in selecting bulls (B). The fitted models included:

Model	Likelihood Ratio χ^2	df	Probability
H_0: [B] [PE]	23.270	3	.000
H_1: [BE] [PE]	8.624	2	.013
H_2: [BE] [BP] [PE]	6.843	1	.009
H_3: [BEP]	0.000	0	1.000

As shown by chi-square statistics and probability levels, only the model with the three-way interaction adequately fits the data. This means that the order of the response alternatives influenced the relationship between the latter two variables.

damental issues is urgently needed if attitude surveys [or evaluations] are to avoid tripping on their own artifacts." In situations where response-order effects are suspected, Payne (1951) recommended using two or more forms in which the response options are alternated. But, the cost of producing multiple forms and bookkeeping efforts can become prohibitive for evaluators who have limited budgets. However, efforts to reduce response-order effects can enhance the validity of evaluation studies.

REFERENCES

BECKER, S.L. (1954). Why an order effect. *Public Opinion Quarterly, 18*, 271-278.

BISHOP, G.F., HIPPLER, H., SCHWARZ, N., & STRACK, F. (1988). A comparison of response effects in self-administered and telephone surveys. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 321-340). New York: John Wiley & Sons.

BROOK, D., & UPTON, G.J.G. (1974). Biases in local government elections due to position on the ballot paper. *Applied Statistics, 23*, 414-419.

CARP, F.M. (1974). Position effects on interview responses. *Journal of Gerontology, 29*, 581-587.

DILLMAN, D.A. (1978). *Mail and telephone surveys: The total design method*. New York: John Wiley & Sons.

HOMANS, G.C. (1961). *Social behavior: Its elementary forms*. New York: Harcourt, Brace and World, Inc.

KALTON, G., COLLINS, M., & BROOK, L. (1978). Experiments in wording opinion questions. *Applied Statistics, 27*, 149-161.

KRAUT, A.I., WOLFSON, A.D., & ROTHENBERG, A. (1975). Some effects of position on opinion survey items. *Journal of Applied Psychology, 60*, 774-776.

KROSNICK, J.A., & ALWIN, D.F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly, 51*, 201-219.

MUELLER, J.E. (1970). Choosing among 133 candidates. *Public Opinion Quarterly, 34*, 395-402.

PAYNE, S.L. (1951). *The art of asking questions*. Princeton, NJ: Princeton University Press.

RUGG, D., & CANTRIL, H. (1944). The wording of questions. In H. Cantril (Ed.), *Gauging public opinion* (pp. 23-50). Princeton, NJ: Princeton University Press.

SCHUMAN, H., & PRESSER, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. Orlando, FL: Academic Press.