

FACULTY WORKING PAPERS
College of Commerce and Business Administration
University of Illinois at Urbana-Champaign
February 11, 1974

EXPERIMENTAL EFFECTS OF
FOLLOW-UP METHODS, QUESTIONNAIRE LENGTH,
AND MARKET HETEROGENEITY IN MAIL SURVEYS

A. Marvin Roscoe, Dorothy Lang, & Jagdish N. Sheth

#158

EXPERIMENTAL EFFECTS OF FOLLOW-UP METHODS, QUESTIONNAIRE LENGTH,
AND MARKET HETEROGENEITY IN MAIL SURVEYS

A. Marvin Roscoe
Dorothy Lang
and
Jagdish N. Sheth

A three factorial 64 cell fixed effects experimental design was created to investigate the impact of four different follow-up methods, two different lengths of the questionnaire and eight distinct markets on the response rate. The telephone reminder was significantly better and the telephone interview without alert was significantly worse in generating returns. The questionnaire length had no effect on the response rate. Finally, the response rate was significantly different across some of the eight markets.

A. Marvin Roscoe is a Marketing Supervisor and Dorothy Lang is a Staff Associate in the Market Research Section of the AT&T Company. Jagdish N. Sheth is I.B.A. Distinguished Professor and Research Professor at the University of Illinois, Urbana-Champaign. The authors wish to express their appreciation to Mr. R. B. Ellis and Mr. N. J. Mammana of AT&T Company and to Mr. Lee J. Cobb of Crossley Surveys, Inc. who performed the field work.

EXPERIMENTAL EFFECTS OF FOLLOW-UP METHODS, QUESTIONNAIRE LENGTH,
AND MARKET HETEROGENEITY IN MAIL SURVEYS

A. Marvin Roscoe
Dorothy Lang
and
Jagdish N. Sheth

In the last three to four decades, considerable progress has been made in the art of data collection and in the reduction of nonsampling error in survey research.¹ However, there are still a number of areas in opinion research about which either very little is known or the existing evidence is inconclusive. This is especially true for mail surveys where there is very little agreement about the factors which affect response rate or the magnitude of their influence although considerable empirical research has been conducted.² For example, the interaction effects of interviewing method or follow-up procedures and questionnaire length remain unknown. We believe this lack of knowledge is due to at least three causes: many variables influencing the response, lack of a comprehensive theory, and limited resources allocated to fundamental research on survey techniques.

There are three aspects of mail surveys which are of considerable practical importance and yet are not well understood. First, what are the effects of different follow-up procedures in mail surveys? Is a post card reminder as

¹For a comprehensive bibliography on research on data collection see D. R. Potter et. al. Questionnaires for Research: An Annotated Bibliography on Design, Construction and Use, Portland, Pacific Northwest Forest and Range Experiment Station, USDA, 1972; and for a comprehensive theory of nonsampling error, See S. Sudman and N. Bradburn, Toward a General Theory of Response Effects in Surveys, Aldine Press (in press).

²Christopher Scott, "Research on Mail Surveys", Journal of the Royal Statistical Society, Vol. 124 (2), 1961, pp. 143-205; Paul L. Erdos, Professional Mail Surveys, New York: McGraw-Hill Book Company, 1970.

good as a telephone reminder? Is it better to switch to a telephone or personal interview for those respondents who do not return the mail questionnaire within a prespecified time period? While there is a general consensus that any follow-up or prodding generates additional response,³ there is little agreement about the relative effects of different follow-up procedures.

Second, what is the effect of questionnaire length on the response rate? It seems logical to expect that response rate should decrease with an increase in questionnaire length, but most of the evidence to date has failed to support this hypothesis. In fact, the opposite result has been found in a number of studies.⁴ Unfortunately, deductive reasoning and face validity are more convincing to most researchers than isolated empirical research, and consequently response rate is still believed to decrease with increased questionnaire length.

The third and one of the most important questions is the determination of interaction effects among questionnaire length, follow-up method, and various respondent characteristics. For example, is one particular follow-up method more effective for a short questionnaire and another technique better for a long questionnaire? Also, are some respondents more receptive to a particular kind of follow-up method or questionnaire length than other respondents?

STUDY DESIGN

Answers to these questions about the effects of questionnaire length, follow-up method, respondent characteristics, and their interactions are critical to

³J. R. Hochstim and D. A. Athanasopoulos, "Personal Follow-up in a Mail Survey: Its Contribution and Its Cost," Public Opinion Quarterly, Vol. 34, 1970, pp. 69-81; and B. K. Eckland, "Effects of Prodding to Increase Mail-Back Returns," Journal of Applied Psychology, Vol. 49, 1965, pp. 165-169.

⁴See Christopher Scott, op. cit., for a listing of studies.

AT&T which periodically sends a mail questionnaire to its residence customer panel consisting of 30,000 members. These customers are asked to participate in a survey by completing and returning a four-page questionnaire, and data are used as input for AT&T's Market Research Information System (MRIS). The 30,000 residence customers in this longitudinal panel have been chosen to be representative of all telephone users at the national and state levels, and for local areas within some of the larger states. To ensure representativeness of the sample in the fixed panel, it is important to maximize the questionnaire response rate to include as many of the panel members as possible in the subsequent analysis.

The four-page questionnaire mailed to panel members requests information on telephones and their location in the home, telephone usage, housing, mobility, demographics, and socioeconomic characteristics. Together with information from company billing records, these data are used to develop models of telephone behavior patterns. The survey procedure uses an alert post card followed by a first mailing of the questionnaire with a cover letter asking the respondent to cooperate in the survey. A reminder post card then is sent followed by a second mailing of the questionnaire to those who did not fill out and return the first questionnaire.⁵

At the start of the present study, three years had passed since the initial customer data had been collected, and it was necessary to obtain more current and up-to-date information by mailing another questionnaire to present panel members. At the same time, it was decided to revise the present four-page

⁵This procedure is detailed in Paul L. Erdos, op. cit.

questionnaire, determine the effects of gathering additional attitudinal information, validate the present survey method, and test alternatives to improve the response rate.

Since past research evidence on the effects of follow-up methods and questionnaire-length on response rate was not decisive in guiding management decisions, a study was initiated to investigate the individual and joint effects of these factors in a number of heterogeneous market areas. The study consisted of a $2 \times 4 \times 8$ three-factorial fixed-effect field experiment and a subsequent set of postexperiment interviews to explore customer attitudes toward the mail survey.

The first experimental factor was questionnaire length. It consisted of two levels: a four-page short questionnaire (slightly revised from the existing questionnaire) and a six-page long questionnaire which included additional attitudinal and perceptual questions about the telephone as a product and as a service. The short questionnaire consisted of 23 items and the long questionnaire consisted of the same 23 items plus 26 attitudinal items. The questionnaire was pretested on a sample of 33 respondents in selected areas for wording, sequence and format.

The second factor was follow-up methods. Four different follow-up methods were selected as feasible alternatives after considerable exploration of the cost and convenience of various follow-up strategies. The four interviewing methods selected used a uniform post card alert and first questionnaire mailing. The follow-up methods varied among experimental treatments and consisted of the present post card and three other methods:

1. Present follow-up method - post card reminder and a second mailing of the questionnaire
2. Second follow-up method - telephone reminder and a second mailing of the questionnaire
3. Third follow-up method - letter alert followed by a telephone interview
4. Fourth follow-up method - telephone interview without alert

The third factor was the market heterogeneity. The national panel in the MRIS system is based on a multistage stratified sample in which the Bell System is geographically divided into one hundred regions, and a sample of 300 residence customers is chosen to represent each region. Based on a hybrid multivariate cluster analysis⁶ of the 100 Bell System markets in terms of socio-economic, demographic and telephone usage characteristics, the following eight geographical areas were chosen as representative market regions: Southeast Massachusetts, San Jose, Arizona, Philadelphia, Eastern Wisconsin, Suburban Chicago, Fort Worth and Alabama. These markets represented eight fixed levels of the third factor in the experimental design.

Within each test market, questionnaires were mailed to a representative random probability sample of 264 customers. In Southeast Massachusetts, 296 questionnaires were mailed to compensate for an anticipated high rate of disconnection of telephone service due to the end of the summer season, and the results were correspondingly adjusted downward. Within each test region, the sample was equally divided among the eight experimental conditions (two questionnaire lengths and the four follow-up procedures) with the result that each of the 64 cells in the three-factorial experiment design had a sample of 33 customers.

⁶A. Marvin Roscoe and J. N. Sheth, "Applications of a Hybrid Cluster Analysis in Industrial and Residential Markets," unpublished working paper, January 1974.

In addition to these three experimental factors, information on respondent characteristics was obtained by using the demographic, socioeconomic, telephone usage, and attitudinal information from the questionnaire itself. These data were used to estimate the respondent attributes which could interact with questionnaire length and follow-up methods to influence reaction to questionnaire length and follow-up methods.

The study began with the 2,144 members of the sample receiving an alert post card on or about Monday, September 18, 1972 and the questionnaire on Wednesday, September 20. The post card and telephone reminder groups received a second questionnaire on Friday, September 29.

The telephone reminders and follow-up telephone interviews were conducted from September 29 through October 2 by the field staff of a national commercial research company in each test region. The mailing dates including those for the follow-up procedures were designed in such a way as to compensate for the regional variations in postal delays. A total of four attempts were made to contact each respondent for the telephone reminders and telephone interviews. In addition, the interviewer was allowed to speak to any responsible adult in the household if the person listed on the telephone billing account was not available. Finally, calls were made at various times of the day and evening which the local supervisor considered most appropriate. In short, every possible effort was made to minimize the differences in situational factors (time, person and opportunity to reach) between the telephone and mail interviews.

The mail questionnaires were accompanied by cover letters on the letterhead of the local Bell Telephone Company. However, the content of the alert post card,

the cover letter and the reminders was identical across all the eight test market regions. All mailings were posted by the survey research manager of each local Bell Telephone Company and the return mail was delivered to that company. First class mail was used for both the delivery and the return of the questionnaire because registered and special delivery mails are not only costly but were found in prior research to be a source of irritation to the respondents.

Post-experiment interviews were conducted with a subsample of respondents including those who did not return the questionnaire by the end of the experiment. These interviews were conducted to determine the respondent's feelings toward the survey in general, and their reactions to the time intervals in the mailing and follow-up procedures, to verify the demographic information provided in the survey, and to check the understanding of specific questions. Interviewing began on October 30 and was completed by November 6. A total of 393 respondents were successfully interviewed and were approximately equally divided among the three experimental factors. One-third of the interviews were conducted in person and the others by telephone.

An analysis of variance was performed on the data from the basic experiment, and the three experimental factors and all combinations of their interactions were tested for significance. In addition, specific questions on the test questionnaire were crosstabulated by questionnaire length, follow-up method, and market. From the post-experiment survey, usage and attitude information were tabulated by whether or not the customer returned the earlier experimental questionnaire, questionnaire length, and follow-up method. The post-experiment survey allowed respondents and non-respondents to be compared in terms of other experimental factors.

RESULTS AND DISCUSSION

The results of the study are summarized in Table 1. The overall response rate was 67.3 percent. However, there are significant differences across various experimental treatments of the three factors.

First, the four follow-up methods produced considerably different response rates across the markets and the two versions of the questionnaire. The telephone reminder was the best follow-up with an average response rate of 76.4 percent, followed by the post card reminder with an average of 69.6 percent, the telephone interview with an alert an average of 65.9 percent, and finally the telephone interview without an alert with an average of 57.0 percent response rate. Based on the Newman-Keuls test on the pairwise differences within an experimental factor,⁷ we found that most of the differences across follow-up methods can be attributed to these two conditions: telephone interview without alert is significantly worse ($p < .01$) than the other three follow-up methods, and the telephone reminder is significantly better ($p < .05$) than the other three follow-up procedures in generating the response rate.

Second, there are no significant differences in the response rate between the short and the long questionnaire contrary to the expectations. Our results thus add to the growing evidence that the length of the questionnaire, within reasonable limits, has virtually no adverse effects on the response rate. In fact, we find a slight although nonsignificant improvement in the response rate with the longer questionnaire.

⁷See B. J. Winer, Statistical Principles in Experimental Design, Second Edition, New York, McGraw-Hill Book Company, 1971, pp. 191-196.

Third, as expected, the eight markets performed differently to the experimental conditions. The best response rate across various experimental conditions came from the Fort Worth area, and the worst response rate came from the Alabama area. Once again, by using the Newman-Keuls test, we found that most of the differences across the eight markets came from these conditions: Alabama had a significantly lower response rate than all other markets ($p < .05$) and Fort Worth had a significantly higher response rate than all other markets ($p < .05$).

In order to quantify the significant differences across the three experimental factors and to isolate the main and the interaction effects, an analysis of variance was performed on the data utilizing the Tukey's test for nonadditivity.⁸ The results are summarized in Table 2. As expected, both the factors of follow-up methods and market heterogeneity had significant main effects and the questionnaire length had no significant main effects. We also found significant interaction effects between the follow-up methods and the market heterogeneity factors and to a much lesser extent between questionnaire length and market heterogeneity.

The post-experimental survey was conducted to assess attitudinal differences among respondents belonging to different experimental conditions. We should expect no significant differences across experimental groups in their general attitudes toward survey participation when both the responders and nonresponders are included in each experimental group. This lack of difference can suggest that the experimental groups were homogeneous in their feelings toward survey

⁸See J. E. Winer, *op. cit.*, pp. 394-397.

participation which can provide additional credence to the analysis of variance results.

Table 3 summarizes the feelings of respondents toward survey participation across different experimental conditions. We have not broken down further by markets due to extremely small sample sizes for many categories of responses coded from open ended answers obtained in the post-experiment survey. As can be seen from Table 3, the percentages of positive, negative and neutral comments are remarkably similar across the experimental conditions when both responders and nonresponders are grouped together in each experimental condition. On the average, 81 percent of the respondents had positive feelings toward the survey. While the degree or manner of positive feelings varied somewhat across experimental groups, generally, there is a remarkable degree of similarity in the data. A total of 15 percent of the respondents had negative feelings toward the survey, and the balance of 7 percent were either neutral or had no opinion. This homogeneity of experimental groups with respect to their general feelings toward survey participation gives stronger credence to the results discussed earlier.

In order to obtain insights into the reactions toward the specific experimental conditions (questionnaire length and follow-up procedures), the respondents were directly asked how they felt about the specific experimental condition to which they were subjected. Table 4 summarizes the feelings toward the length of the questionnaire for each of the two levels of the experimental condition. It should be noted that only those who cooperated in the survey (returned the questionnaire) are included in this analysis. As can be seen from the table, there are no real differences between the two groups of short and long questionnaires

although there is a slightly more incidence of positive comments in the longer questionnaire group. More interestingly, the group with longer questionnaire considered it to be "comprehensive/well designed" significantly more than the group with the shorter questionnaire. However, this is somewhat offset by a higher percentage in the shorter questionnaire group with respect to "easy to fill out/not too long" and "questions were to help the company improve service" categories of positive responses. With respect to the negative comments, it is surprising to find that considerably more respondents in the shorter questionnaire group felt that some of the questions were too personal. As would be expected, slightly more number of respondents in the longer questionnaire group felt that some of the questions were complicated.

Table 5 summarizes the post-experiment comments specifically related to the follow-up methods. Due to the small sample size, we have combined respondents for both telephone interview follow-up procedures into a single category. There are some interesting and significant differences across the three groups between in the percentages of positive comments and the specific types of positive comments. Significantly more number of respondents in the telephone interview group gave positive comments than in the telephone reminder and the post card reminder procedures (87 percent vs. 51 and 47 percent, respectively). This is somewhat surprising in view of the fact that the response rate was lower in the telephone interview follow-up methods. However, the bulk of the positive comments is more in the nature of doing a favor or ingratiating the telephone interviewer than for the advantage of the follow-up method. Thus, 41 percent stated that they didn't mind the telephone interview and 30 percent stated that they were glad to help the company. In contrast, only 10 percent felt that phone call made the survey clearer and 27 percent favored telephone interview over the

mail questionnaire. Examining the positive comments in the telephone and post card reminder groups, however, we find that most of the positive comments are directly anchored to the method of follow-up procedures. Many respondents considered the specific follow-up procedure a good idea or a good reminder.

Turning toward the negative comments, we find that some of the cooperative respondents in the telephone interview group didn't like telephone interviews or resented the telephone survey when they had an unlisted number. Surprisingly, some of the respondents in the telephone reminder procedure found it to be "annoying or pushing me."

Finally, there is a large percentage of respondents who claimed that they didn't receive the post card reminder or didn't remember receiving it (39 percent). Similarly, some of the respondents in the telephone reminder group also claimed not receiving the reminder or not remembering it. We believe that these are excuses for not fully participating in the survey. Furthermore, it is easier for the respondent to claim nonreceipt of the reminder if it is through the mail than if it is through the telephone. In fact, the superiority of the telephone reminder found in the study may be solely due to this factor.

Our analysis of the eight distinct markets produced no significant differences except that the respondents in the Fort Worth area had significantly higher income than the 1970 Census Tracts Data for the area. Due to a very high development of multiple telephone numbers in the upper income households in the Fort Worth area, the random probability sample had generated more numbers of upper income respondents. The lack of systematic differences in the feelings toward survey participation or toward specific experimental conditions across the eight

markets is disappointing as we strongly believe that we need to consider the concept of market segmentation in survey research: some questioning procedures are better suited for one segment of the total respondents and others are better suited for other segments. No single technique of data collection can fully satisfy all segments of the population.

SUMMARY AND CONCLUSIONS

This study was designed to test and evaluate two specific hypotheses for management. First, could an instrument of 26 attitude items be added to the present questionnaire without deleterious effects on response rate? Second, is there a practical alternative to the present follow-up procedure that would increase overall response rate? The results clearly indicated that the attitude questions can be added to the questionnaire without any measurable effect in response rate and that this revised larger questionnaire would not influence the selection of the follow-up procedure. The telephone reminder with the second mail questionnaire is clearly the best overall strategy to replace the present post card reminder. The telephone reminder was consistently higher over all markets and made a more positive and memorable impression on the survey participants. The emphasis here was placed on increased response rate which would result in improved panel validity and effectiveness; therefore, no direct cost comparisons were made. However, since the survey mailing and collection are done locally, the telephone costs compare favorably with the cost to print, address and mail the reminder post cards.

Since the telephone reminder and the post card reminder are similar in nature and differ only in form, it is possible to suggest that if specific markets find that the results with the telephone reminder do not achieve a better return

than their previous experience, they should continue to use the post card reminder. This tailoring of the survey procedure to the specific markets suggests that the best survey methodology might be different for various market segments. In particular, the results suggest that high socioeconomic customer groups and those in the Southwest respond best to the alert and telephone interview. However, this procedure should not be used in the Southern areas where the best response was obtained by a telephone reminder. The telephone interview is also not recommended as of follow-up procedures in and around the major metropolitan areas where the telephone reminder was preferred in the city and the post card reminder in the suburban areas. Additional research, however, is necessary to confirm and extend these findings.

TABLE 1.

Questionnaire Length	Short Questionnaire					Long Questionnaire				
	Follow-Up Methods					Follow-Up Methods				
	Post Card Reminder	Telephone Interview	Telephone Interview	Alert & Telephone Interview	Average Response Rate	Post Card Reminder	Telephone Interview	Telephone Interview	Alert & Telephone Interview	Average Response Rate
Heterogeneity										
Alabama	43.3	67.7	56.2	32.3	53.2	70.0	68.8	39.4	48.4	53.2
Southeast Massachusetts	70.3	70.3	54.1	64.9	64.2	59.5	78.4	58.4	58.4	64.2
Suburban Chicago	74.2	87.5	50.0	71.9	53.5	75.0	63.6	31.2	54.6	53.5
Ft. Worth	75.8	71.9	71.9	87.1	77.6	81.8	81.8	62.5	87.9	77.6
Eastern Wisconsin	69.7	87.9	57.6	59.4	68.8	72.7	87.9	51.5	63.6	68.8
Philadelphia	57.6	75.0	54.6	69.7	65.1	48.5	69.7	66.7	78.8	65.1
Arizona	78.8	66.7	57.6	69.7	74.9	75.0	84.9	78.8	87.9	74.9
San Jose	80.0	84.4	65.6	62.5	70.7	81.8	76.8	56.2	58.1	70.7
Average Response Rate	68.7	76.4	58.4	64.7	67.3	70.5	76.5	55.6	67.2	67.3

TABLE 2
ANALYSIS OF VARIANCE OF QUESTIONNAIRE RETURNS*

<u>Source of Variation</u>	<u>Sum of Squares</u>	<u>Degrees of Freedom</u>	<u>F Ratio</u>	<u>Significance</u>
Questionnaire Length	0.26	1	0.04	N.S.
Follow-up Methods	342.53	3	16.94	.001
Markets	350.93	7	7.44	.001
Questionnaire Length X Follow-up	7.51	3	0.37	N.S.
Questionnaire Length X Market	103.70	7	2.20	.10
Follow-up X Market	311.87	21	2.20	.05
Questionnaire Length X Follow-up X Market (Error)	141.58	21		

*The Anova procedure is based on B. J. Winer, Statistical Principles in Experimental Design, 2nd Edition, New York: McGraw-Hill Book Company, 1971, pp. 321-327 and 384-387. The use of the three factor interaction for the estimate of the error variance was based on Tukey's test for non-additivity which was rejected as not significant.

TABLE 3
ATTITUDES TOWARD SURVEY PARTICIPATION*

	Questionnaire Length		Follow-up Method			
	Long	Short	Telephone Reminder	Post Card Reminder	Telephone Interview No Alert	Telephone Interview with Alert
	(n=192)	(n=194)	(n=97)	(n=97)	(n=96)	(n=96)
<u>Respondents Making a Positive Comment</u>	81%	80%	82%	81%	80%	79%
Happy to do it/Glad to help/An honor	27	18	25	20	26	20
Didn't mind/Was OK	43	43	35	47	45	46
Gave chance to express feelings/Helping to improve service/To my benefit	11	15	17	12	11	12
Like Phone Co/Happy they were interested/Good way to find out about service	15	12	20	10	17	9
Questions were easy	3	2	0	1	4	5
<u>Respondents Making a Neutral Comment</u>	7	8	7	5	7	9
<u>Respondents Making a Negative Comment</u>	15	16	13	16	14	18
Don't like Phone Co/ Unhappy with service/ Survey not worthwhile	3	3	2	2	4	4
Questions too personal/ Too long/Confusing	8	9	10	5	8	8
Didn't have time/ An imposition	5	9	5	6	4	9

*Based on all 386 respondents to the post-experiment survey including those not returning the test questionnaire.

TABLE 4
ATTITUDES TOWARD THE QUESTIONNAIRE*

	Long Questionnaire (n=159)	Short Questionnaire (n=161)
<u>Respondents Making a Positive Comment</u>	70%	67%
Was good/nice/fine	26	19
Didn't mind it	15	16
Comprehensive/Well designed	28	17
Gave opportunity to express feelings	11	9
Easy to fill out/Not too long	10	16
Questions were to help the Company improve service	6	11
<u>Respondents Making a Neutral Comment</u>	9	7
<u>Respondents Making a Negative Comment</u>	30	30
Some questions too personal	4	14
Some questions complicated	12	8
Some questions inappropriate/ Redundant/Foolish	10	9
All other negative comments	8	6

*Based on the 320 respondents to the post-experiment survey who had completed the test questionnaire.

TABLE 5
ATTITUDES ABOUT FOLLOW-UP METHODS

	Telephone Reminder (n=97)	Post Card Reminder (n=97)	Telephone Interviews* (n=41)
<u>Respondents Making a Positive Comment</u>	51%	47%	87%
Good idea/Good reminder	23	22	NA
Made it seem important to respond	8	9	0
Was fine/Didn't mind	0	30	41
Showed they care	6	0	0
Rather telephone interview than questionnaire	NA	NA	27
Phone call made survey clearer	NA	NA	10
Caller was courteous, pleasant	23	NA	17
Glad to help	0	0	30
<u>Respondents Making a Negative Comment</u>	11	7	18
Annoyed/Pushing me	11	5	0
Don't like telephone interviews	NA	NA	15
Have unlisted number resent the call	0	NA	7
<u>Don't Remember/Didn't Receive</u>	11	39	0

*Alert group and no alert group combined. Includes only respondents given telephone interview.