MAIL SURVEY RESPONSE RATE
A META-ANALYSIS OF SELECTED TECHNIQUES FOR INDUCING RESPONSE

RICHARD J. FOX
MELVIN R. CRASK AND
JONGHOON KIM

Abstract This article reports the results of a meta-analysis of experimental studies which have examined ten different factors felt to influence response rates to mail surveys. The form of meta-analysis used clearly defines the individual impact of each of the factors examined. Results indicate that prenotification and follow-ups increase the response rate, as does the type of outgoing postage used. Furthermore, studies sponsored by a university receive greater returns. Increases in the size of the monetary incentive used appear to have decreasing marginal gains. Finally, some evidence exists to suggest that the color of the questionnaire influences response rate.

The extensive use of mail surveys coupled with the low response rates typically encountered has made the issue of mail survey response rate improvement an intriguing topic. High response rates have the obvious benefits of increased sample size, reduced costs associated with follow-up contacts, and reduced concern over nonresponse bias. Many experiments have been conducted to measure the effectiveness of one or more of the factors expected to influence response rates.

The purpose of this paper is to apply a form of meta-analysis to reported experimental findings to isolate and estimate the individual impact of several typical response rate enhancement techniques. The meta-analysis technique used avoids the problems associated with past similar efforts but is more restrictive in the factors which can be examined.

RICHARD J. FOX and MELVIN R. CRASK are Associate Professors of Marketing at the University of Georgia. JONGHOON KIM is a Ph.D. candidate in the Marketing Department of the University of Georgia.

Public Opinion Quarterly Volume 52:467–491 © 1988 by the American Association for Public Opinion Research
Published by The University of Chicago Press / 0033-362X/88/0052-04/$2.50

Background

Empirical studies concerned with increasing response rates to mail surveys abound, but most such studies examine the impact of only one or two response rate factors. Several excellent qualitative literature reviews of these studies have appeared (see Kanuk and Berenson, 1975; Linsky, 1975). However, inconsistent findings severely limit the ability of the review authors to make generalizations in such reviews. The difficulties encountered are best stated by the author of one of these reveiws: "Given the substantial body of research on the mail questionnaire, much of it experimental, it seems surprising that more has not been learned. Findings are inconsistent for many of the techniques commonly recommended in research texts" (Linsky, 1975:100).

Part of the problem experienced in traditional qualitative literature reviews stems from the fact that differences in sample size, sample composition, and experimental executions are often ignored. Studies tend to be merely categorized as concluding that specified techniques improve or do not improve response rates. No formal assessment of the size of the respone rate effect achieved by any factor is included.

Recently, quantitative reviews of past findings have begun to appear. These fall into the category of meta-analysis, which broadly refers to statistical techniques for quantitatively aggregating findings from individual studies. For example, Heberlein and Baumgartner (1978) were able to explain 51% of the variance in final response rate with two variables: salience of the topic to the respondent and number of contacts. Additional variance was explained by sample characteristics, number of questions asked, type of postage, and the use of a special third contact. Eichner and Habermehl (1981) attempted to replicate those results in a sample of German and Austrian studies with little success. Goyder (1982), on the other hand, was able to replicate the findings more successfully using another sample of U.S. studies.

Meta-Analysis Techniques

In a meta-analysis, the findings of each study are treated as independent observations which may be combined to calculate an overall, or "average," effect. "By recording the properties of studies and their findings in quantitative terms, the meta-analysis of research invites one who would integrate numerous and diverse findings to apply the full power of statistical methods to the task" (Glass, McGaw, and Smith, 1981:21). When combining study results, one must try to insure that each study is as near a replication of the others as possible. Otherwise, the unique aspects of each particular study enter into the estimate of

the average effect size. Unless the impact of these unique components can be extracted, the overall effect estimate can be misleading.

The meta-analysis technique used in the studies mentioned above is regression analysis (ordinary least squares). The dependent variable is the observed response rate and the predictor variables describe the survey conditions and executional details. Four problems are of particular concern when using a regression approach.

The first problem with the regression approach is missing data. Relevant information about how the study was conducted is frequently omitted from articles, so the value of one or more predictor variables is often unknown for an observation. Judgments must be made regarding what values to assign predictor variables in these instances to avoid eliminating an excessive number of variables from the analysis, or authors must be contacted to provide all of the necessary details. As Heberlein and Baumgartner (1978) and Eichner and Habermehl (1981) experienced, contacting authors and requesting relevant information only partially alleviates the problem.

A second problem is the subjectivity which enters into the assignment of values to nonquantitative predictor variables. For example, Heberlein and Baumgartner (1978) included salience of the questionnaire topic to the survey audience as a predictor variable. A scale measure of salience was subjectively determined by the authors. Goyder (1982:551) found disagreement between some of his results and those of Heberlein and Baumgartner (1978) and stated that "disagreement between the data sets is greatest for variables where judgement entered into coding decisions."

The third problem is that individual study results are typically weighted equally in the analysis. The results of a study conducted with a mailout of 50 are treated the same as the results of a study with a mailout of 500. A more sophisticated technique (generalized least squares) is required to accommodate such differences.

Finally, regression analysis is also subject to multicollinearity among the predictor variables, which causes instability of the estimates of the regression coefficients. Although the estimated regression equation may be fine from a prediction perspective, interpretations regarding the impact of individual predictor variables are questionable in the presence of multicollinearity. Eichner and Habermehl (1981) reported substantial multicollinearity among the predictor variables in a regression analysis of response rates based on a sample of German and Austrian mail survey studies. Moreover, they found that their results differed considerably from those of Heberlein and Baumgartner (1978), and suggested that multicollinearity and cultural differences were likely reasons. In their reply, Heberlein and Baumgartner (1981) addressed the multicollinearity issue. They report average multiple correlation

coefficients of about .5 for the prediction models reported. Further, using a statistical test of hypothesis procedure, the hypothesis of a singular correlation matrix for the predictor variables is rejected for one model, but not for another. They also report that symptoms of severe multicollinearity are not present in their data, but acknowledge the possibility of problems with the stability of their estimates of the regression coefficients. Goyder (1982) applied a similar analysis to a somewhat different data base and reported comparable levels of multicollinearity and mixed results in terms of reproducing the Heberlein and Baumgartner (1978) findings. In sum, the issue of the impact of multicollinearity in the regression estimates of the influence of various factors on response rate is unresolved.

A different approach to meta-analysis is to accumulate statistical evidence against a null hypothesis across individual studies to determine an overall level of significance at which the null hypothesis can be rejected. Yu and Cooper (1983) used this approach to investigate response enhancement techniques by combining findings across a sample of personal, mail, and telephone interviewing surveys. Armstrong and Lusk (1987) used a similar approach to investigate the response rate effect of various forms of return postage.

Accumulating statistical evidence from individual findings addresses the issue of whether the aggregated information is indicative of a statistically significant effect. However, the problem of developing a combined estimate of the effect remains. Armstrong and Lusk (1987) used the average effect across the studies included in the analysis as the estimate of effect size in their analysis pertaining to return postage alternatives. Yu and Cooper (1983) reported averages of response rates, weighted by sample size, in their study encompassing personal, telephone, and mail surveys.

In this paper a meta-analysis approach similar to that of Armstrong and Lusk (1987) and Yu and Cooper (1983) is used to accumulate statistical evidence and to isolate and estimate the effect sizes of various typical response rate enhancement techniques in the context of mail surveys only. A "treatment versus control" approach is adopted to eliminate the impact of unreported factors common to both the treatment and control groups in each experiment (see Armstrong and Lusk, 1987). Hence, unknown executional details do not present a problem. Also, only response rate variables which do not entail subjective coding are investigated.

This study complements instead of replicating previous research efforts. First, some of the response rate enhancement techniques examined here have not been examined in previous quantitative summaries. Second, while the results of this study do not include an equation for predicting response rate as has been done previously, the estimates of

the response rate effects of individual response enhancement techniques are not subject to problems created by multicollinearity. Further, the effect size estimates are "optimal" aggregate estimates of effect size in that the estimation method takes into consideration the variances of the individual study effects. These variances depend on the treatment and control group sample size and response rates.

In the next section, the procedure used to locate the studies included in the meta-analysis is described. Then the nature of the studies reviewed is discussed, including such things as where and when they appeared and the scope of factors which have been examined. The following two sections detail the meta-analysis techniques used and the results of these analyses. The article concludes with a discussion of the implications of the findings for survey researchers.

Literature Search Procedure

The treatment versus control approach dictates restricting attention to studies whose primary thrust was to experimentally investigate the impact of enhancement techniques on mail survey response rate (see Mazzuca, 1982; Armstrong and Lusk, 1987). A census of such experimental studies was attempted. Obviously, all printed material cannot be examined for relevance, and some procedure must be used to identify sources. The procedure used was an iterative one which is common to many searches (see Churchill et al., 1985). First, a set of three "seed" journals (the *Journal of Marketing Research, Public Opinion Quarterly,* and the *Journal of Applied Psychology*) was selected to initiate the search. These journals were chosen because they were known to have published a large number of articles dealing with response rates. Individual issues of each of these journals for the last 25 years, or from the date of inception if the journal had been in existence less than 25 years, were examined. The search was expanded beyond these seed journals in two ways. A computerized keyword literature search was conducted, and all articles not previously identified as relevant were examined. The computerized search was conducted using PSYC/INFO, which is comparable to the information included in *Psychological Abstracts* but also includes dissertations. Also, the reference list from each relevant article located was used to identify additional sources. The reference list of each new source was scanned in the same manner to locate any new references. This reference list scanning procedure was continued until no new sources were discovered. The final reference list thus approximates a census of published materials on mail survey response rate effects for the previous 25-year period (see Churchill et al., 1985).

One of the concerns when conducting a meta-analysis is the "file-drawer problem," a potential bias due to the fact that the analysis is entirely or primarily based on results reported in published literature (Rosenthal, 1984:107–108). Published studies tend to be ones where statistically significant relationships have been found (Wolf, 1986). Accessing unpublished materials is very difficult at best; yet, omission of the results of unpublished studies may bias the estimates obtained from the meta-analysis if the omitted studies are more likely to include nonsignificant results.

Fortunately, this potential bias may be less severe in this study than in many meta-analysis studies. The studies examined in this research were experimental in nature, and very often were multifactor. Consequently, nonsignificant results for a single factor were frequently reported. Of the 214 individual estimates of effects accumulated in this research, only 23% are statistically significant. The number of reported nonsignificant findings would seem much larger than is typically the case for academic research. Therefore the extent of any file-drawer bias should be smaller in this study than one may typically encounter.

Overview of Search Findings

The search identified 148 articles whose titles suggested that the article potentially dealt with mail survey response rate factors. As might be expected, the bulk of these articles appeared in sociology and marketing journals (65 and 45, respectively). The 148 articles were reduced to 82 articles where one or more response rate factors had been experimentally manipulated (a list of these 82 articles is included in the Appendix); 56% of these experiments were published in the 1970s, as compared to only 17% in the prior decade. The remainder of the studies have been published since 1980. Later studies tended to involve more sophisticated designs, encompassing multiple factors and/or multiple treatment levels.

A vast array of specific response rate variables were manipulated in the 82 articles uncovered. These variables have been summarized into the five broad response rate issues shown in Table 1. By far the most frequently examined issue has been the cover letter which accompanies the mail survey. Studies have looked at the effect of various appeals to the respondent, whether to personalize the letter, whether to specifically state the confidentiality conditions, and whether to indicate a return deadline for inclusion in the study.

The next most commonly encountered issue was the use of an incentive. Studies have explored not only the effect of the amount of the incentive but also the effect of whether the incentive is enclosed or

Table 1. Summary of Response Rate Issues Explored by Previous Studies

Response Rate Issues	No. of Comparisons[a]	Variables Examined[b]
Cover letter	62	Personalization, appeal used, *use of postscript, notification of cutoff date,* form of letter, *sponsorship,* statement of confidentiality/anonymity
Incentives	30	*Amount,* enclosed/promised, type, incentive recipient (e.g., charity)
Respondent contact	29	*Use of prenotification,* best form of prenotification, use of follow-ups, proper number of follow-ups, timing of follow-ups, best form of follow-ups, home vs. work address
Postage/mailing	25	*Type of outgoing postage, type of return postage,* type of envelope
Questionnaire	19	Topic, length, *color,* format

[a] The sum of these figures is larger than the number of articles examined because many articles were multifactor.
[b] Italicized variables included in meta-analysis.

promised. The use of noncash incentives and the promising of gifts to charity have also been examined.

Respondent contact issues (how, when, and how often to contact potential respondents) and the type of postage or envelope to use have been examined nearly as often as has the use of incentives.

Interestingly, the least often examined issue is the survey instrument itself. Most studies which have focused upon the questionnaire examined the effect of questionnaire length on response rate. A few have looked at other variables such as the impact of the color or the format of the questionnaire.

The specific manner in which some variables have been operationalized varies tremendously from study to study. For example, when studying the effect of cover letter personalization, researchers have personalized the salutation, the body of the letter, the signature, and/or the postscript. The same is true for questionnaire length, confidential-

ity, noncash incentives, and the type of appeal used in the cover letter. We felt that the way some factors were operationalized varied too greatly from study to study to provide reasonable overall estimates of a common effect when aggregated. This precluded analyzing some factors, such as personalization and questionnaire length, which are of interest to both practitioners and researchers. However, 9 response rate factors were identified for which multiple studies existed, and for which the effect being investigated was the same across studies: prenotification by letter, follow-up postcard, first-class outgoing postage vs. bulk-rate, stamped outgoing postage vs. metered, stamped return postage vs. business reply, notification of a cutoff date, university sponsorship vs. business sponsorship, a green vs. a white questionnaire, and the inclusion of a postscript asking for cooperation. In addition, enough observations were available to determine the relationship between response rate and the amount of cash incentive (from 10¢ to $1) enclosed with the questionnaire. Of the original 82 experimental articles, only 40 provided data for this study. Those 40 articles are identified by an asterisk in the Appendix.

PRENOTIFICATION BY LETTER

Conceptually, one would expect prenotification by letter to increase the response rate. One reason would be that prenotification alerts people that the survey is coming, thus reducing the likelihood of an interested recipient inadvertently discarding it. A second reason would be that prenotification could establish legitimacy of the survey. Such legitimacy would be a form of trust under the theory of social exchange posited by Dillman (1978).

Twenty-two treatment versus control comparisons were reported in the 8 studies located in which prenotification was manipulated (Ford, 1968; Heaton, 1965; Jones and Lang, 1980; Myers and Haug, 1969; Parsons and Medford, 1972; Pucel, Nelson, and Wheeler, 1971; Stafford, 1966; and Walker and Burdick, 1977). In 19 of the 22 comparisons, prenotification increased response rate. The effect of prenotification ranged from a 9% decrease in response rate to a 47.4% increase. We anticipated that the meta-analysis would reveal that prenotification produces a significant increase in response rate.

POSTCARD FOLLOW-UP

A postcard follow-up would be expected to increase response rates for two reasons. First, the appreciation expressed in the follow-up would be a form of reward in the social exchange process (Dillman, 1978). Second, the follow-up serves to remind nonrespondents that they have

forgotten to complete the survey, a common reason for nonresponse (Dillman, 1978).

Six articles were located which examined the effect of a follow-up postcard (Cox, Anderson, and Fulcher, 1974; Hinrichs, 1975; Jones and Lang, 1980; Peterson, 1975; Watson, 1965; and Wiseman, 1973). In 33 of the 36 comparisons, follow-up increased response rate. The size of the effect ranged from an 11% decrease to a 35% increase when a follow-up postcard was used. We expected the meta-analysis to reveal an overall significant increase in response rate when a follow-up postcard is used.

OUTGOING POSTAGE

Enough studies were found to examine two aspects of outgoing postage: whether first-class postage works better than bulk-rate and whether stamps work better than metered mail. Two reasons exist to suggest that first-class postage would generate higher response rates than would bulk-rate postage. First, bulk-rate postage is associated with "junk mail" and would tend to suggest that the survey is not important. Second, bulk-rate postage does not receive the handling priority of first-class (Dillman, 1978). Stamps, rather than metered mail, might be expected to increase response rates because they are perceived as more personal (Armstrong and Lusk, 1987).

Five studies examined the effect of first-class versus bulk-rate outgoing postage (Brook, 1978; Gullahorn and Gullahorn, 1963; Kernan, 1971; McCrohan and Lowe, 1981; and Watson, 1965). Thirteen of the 18 comparisons found in these five studies revealed an increase in response rate associated with first-class postage. The effect of first-class postage ranged from a 9% decrease in response rate to an increase of 10%. We felt that the meta-analysis of the results would reveal a significant increase in response rate associated with the use of first-class outgoing postage.

Four studies manipulated outgoing postage by using metered mail instead of stamps (Dillman, 1972; Kernan, 1971; Peterson, 1975; and Vocino, 1977). Nine of the 20 comparisons indicated no increase in response rates due to stamped outgoing mail. The size of the effect ranged between a 10% decrease and a 14.8% increase when stamps were used as opposed to metered mail. Nevertheless, we still tentatively expected stamps to outperform metered mail on average.

STAMPED VERSUS BUSINESS REPLY RETURN POSTAGE

Business reply envelopes have the obvious advantages of being less time-consuming for the respondent to prepare for mailing and requiring

postage payment by the sponsor only when returned. However, they may be seen as less personal and thus reduce the return rate (Armstrong and Lusk, 1987). Nine studies were found which examined stamped versus business reply return postage (Brook, 1978; Gullahorn and Gullahorn, 1963; Harris and Guffey, 1978; Jones and Linda, 1978; Kimball, 1961; McCrohan and Lowe, 1981; Peterson, 1975; Watson, 1965; and Wiseman, 1973). In 42 of the 50 comparisons, stamped returns generated a higher return rate than did business reply. The effect of stamps ranged from a decrease in the response rate of 4.5% to an increase of 32% versus business reply. We expected that the meta-analysis would indicate a significant overall positive effect on response rate when stamps were used instead of business reply returns.

NOTIFICATION OF CUTOFF DATE

Because respondents often set a survey aside and forget to complete it at a later date, some researchers have felt that indicating in the cover letter that returns must be received by a certain date might increase the response rate. Five studies were located which tested this hypothesis (Futrell and Hise, 1982; Henley, 1976; Pressley, 1979; Roberts, McCrory, and Forthofer, 1978; and Vocino, 1977). In 9 of the 12 experiments, notification of a cutoff date increased the response rate. The effect ranged from a 13.5% decrease in response rate to a 7.8% increase. We expected a significant overall positive effect for notification of a cutoff date.

UNIVERSITY SPONSORSHIP

Dillman (1978:16) suggests that university sponsorship of the survey may increase the response rate because of the past benefits that the respondents may have received from the university. Four studies were found which examined the effect of university sponsorship (Houston and Nevin, 1977; Jones and Lang, 1980; Jones and Linda, 1978; and Peterson, 1975), and these provided 41 observations. (One study, a $2 \times \times 2 \times 2 \times 2 \times 2$ factorial experiment, provided 16 comparisons, and another, a $2 \times 2 \times 3 \times 2$ factorial experiment, provided 12 comparisons.) University sponsorship failed to increase the response rate in only 4 of the 41 experiments. We anticipated a significant overall positive effect on response rate would be found in the meta-analysis.

COLOR OF THE QUESTIONNAIRE

Three studies were located which examined the impact of using a colored questionnaire instead of a white one (Gullahorn and Gullahorn,

1963; Pressley and Tullar, 1977; and Pucel, Nelson, and Wheeler, 1971). The intent was to determine whether the attention-grabbing effect of a colored questionnaire as opposed to a white one would stimulate response rate. Almost all reported results entailed a green versus a white questionnaire, so the meta-analysis became a study of the aggregate effect of a green versus a white questionnaire.

In eight of the ten experiments reported, the response rate was higher for the green questionnaire. The range of effects varied from a decrease in response rates of 5.6% when a green questionnaire was used to an increase of 9.1%. We expected the meta-analysis to reveal an overall significant increase in response rate associated with a green questionnaire.

POSTSCRIPT ASKING FOR COOPERATION

Handwritten postscripts asking for cooperation could be expected to increase response rates because of the personalization effect. Two studies examined the effect of such a postscript (Childers, Pride, and Ferrell, 1980; Pressley, 1979). Contrary to commonly held opinions, in five of the six comparisons a postscript requesting cooperation reduced the response rate. The effect ranged from an 11% decrease to a 6% increase when postscripts were used. Thus, no prediction could be made in this case.

MONETARY INCENTIVE

Few would question the assertion that a monetary incentive enclosed with the questionnaire will increase the response rate. Aside from the monetary value, Dillman (1978:16) suggests that the enclosed money establishes a trust necessary for the social exchange to occur. Fifteen studies which examined the impact of a monetary incentive on response rate were located (Friedman and Augustine, 1979; Furse and Stewart, 1982; Furse, Stewart, and Rados, 1981; Goodstadt et al., 1977; Hackler and Bourgette, 1973; Hansen, 1980; Huck and Gleason, 1974; Kimball, 1961; McDaniel and Jackson, 1984; Newman, 1962; Pressley and Tullar, 1977; Robin and Walters, 1976; Watson, 1965; Wiseman, 1973; and Wotruba, 1966). In all but two of the 30 experiments reported, incentives increased the response rate.

Armstrong (1975) used ordinary (unweighted) least squares regression analysis to develop a diminishing returns model for the percentage decrease in the rate of nonresponse as a function of the incentive amount. Dillman (1978:16) points out that "the closer the monetary incentive comes to the value of the service performed, the more the transaction tends to move into the realm of economic exchange and the

easier it becomes for many people to refuse it." Incentives are typically well below the perceived value of the service, so it seems reasonable to assume that increasing the incentive reduces the amount by which the perceived value exceeds the incentive. Because of these previously reported relationships, a diminishing returns model was fit to the response effect versus amount of incentive data.

Analysis Procedure

Since the studies included in the analyses were experimental, the authors very clearly explained how the variables of interest were manipulated. In addition, a control group, which received a "zero" level of the treatment (for example, no prenotification, no incentive, etc.) was always included. Although the authors would generally not give detailed descriptions of experimental conditions other than those being manipulated, these conditions were common to all of the groups in the experiment. The impact of these extraneous factors is eliminated by using the difference between the treatment group response rate and the corresponding control group response rate as the measurement of effect rather than just the treatment group response rate. This is a common approach to the meta-analysis (Glass, McGaw, and Smith, 1981; Wolf, 1986; and Armstrong and Lusk, 1987), and the obtained difference is referred to as the effect size. This difference is a "pure" measure of the effect of the factor of interest under the condition that there are no interactions between the factor present in the treatment and the common conditions. Further, as noted by Armstrong and Lusk (1987), previous research has indicated a low likelihood of substantial interaction effects.

For each factor except incentive, the analysis proceeded as follows: The relevant studies were aggregated. A combined significance was obtained by first calculating the significances associated with the individual tests of the hypothesis of no effect from each study (simple z-tests based on the sample response rates for the treatment and control groups). These were combined in the conventional manner (Wolf, 1986) to produce an aggregate significance via the chi-square statistic, $-2\Sigma_{i=1}^{n} Ln\ (\alpha_i) = \chi^2$, with $2n$ degrees of freedom where α_i is the significance or "p-value" of the ith test (Fisher, 1950).

Although the chi-square test described above yields an overall significance level of past research, the average response rate effect size must be estimated separately (Wolf, 1986). A simple average of the observed effect sizes is not the best estimate. The variance, V, of any measurement of the effect is the sum of the variances of the response

rates for the treatment and control groups, which depend on their respective response levels and sample sizes. That is,

$$V = P_T (1 - P_T)/n_T + P_C (1 - P_C)/n_C,$$

where P_T, P_C and n_T, n_C are the response rates and sample sizes for the treatment and control groups, respectively. Since the response levels and sample sizes for the treatment and control groups varied from observation to observation, the variances of the corresponding estimates of the effect, that is, the variances of the differences between response rate for the treatment group and response rate for the control group, are different. The minimum variance unbiased estimate is the weighted average of the individual effects, where the weights sum to unity and are inversely proportional to the respective variances (Lentner and Bishop, 1986). An approximation of the minimum variance unbiased estimate was obtained by using the estimates, derived from the sample data, of the respective variances of the observed effects as the weights in the averaging. The significance of the combined estimate in testing the null hypothesis that the effect is zero was calculated. A simple z-test, based on the ratio of the approximate minimum variance estimate of effect size to the estimate of its standard deviation, was used.

In the case of an incentive, four different treatment levels, that is, amounts of incentive offered, were observed across studies. Hence, a different approach was taken. Response rate effect size was calculated for each study in the same manner as the other factors. The individual effects, and other relevant information, are shown in Table 2. Preliminary examination of the graph of the effect size as a function of the amount of the incentive indicated that a curvilinear relationship existed, and that the model equating the expected effect size to a constant multiple of the square root of the incentive was reasonable. No intercept was included because the expected effect is zero if no incentive is used, that is, the expected values of the treatment and control group response rates are the same. Formally, the mathematical model used is

$$y = \beta\sqrt{x} + \epsilon,$$

where y is the measured effect, x is the amount of incentive measured in cents, β is an unknown parameter, and ϵ is a normally distributed random variable having zero mean; the random errors (ϵ's) are assumed to be mutually independent.

As discussed earlier, the variances of the observations depend on respective response levels and sample sizes and thus differ. Hence, weighted least squares, where the weights are the inverses of the estimated variances of the observations (measurements of effect), was used to fit the model. A second approach to the model was performed

Table 2. Incentive Data

Amount of Incentive (x)	Sample Size Control	Sample Size Treatment	Response Rate Control	Response Rate Treatment	Effect(y)
$ .25	150	150	.247	.373	.126
.25	150	150	.193	.380	.187
.50	100	100	.540	.680	.140
1.00	100	100	.540	.760	.220
.50	100	100	.560	.710	.150
1.00	100	100	.560	.780	.220
.25	604	604	.621	.747	.126
1.00	109	109	.530	.830	.300
.25	811	810	.141	.380	.239
.10	500	500	.348	.446	.098
.10	500	500	.260	.382	.122
.25	1,501	750	.235	.372	.137
.25	75	75	.293	.467	.174

1.00	75	75	.293	.533	.240
.10	136	136	.333	.445	.112
.10	189	199	.191	.462	.271
.10	98	99	.204	.293	.089
.10	8,999	501	.311	.399	.088
.25	8,999	500	.311	.478	.167
.10	29	29	.241	.448	.207
.10	29	29	.414	.586	.172
.10	29	29	.345	.345	.000
.10	29	29	.621	.655	.034
.10	29	29	.276	.414	.138
.10	29	29	.414	.310	−.104
.10	29	29	.414	.586	.172
.10	29	29	.448	.483	.035
.25	50	50	.180	.400	.220
.50	147	147	.388	.518	.129
.25	50	50	.710	.940	.230

by adjusting the incentive based on the consumer price index so that the amount of the incentive could be viewed in constant dollars. However, no improvement in the fit resulted, and only the results based on the unadjusted incentive amounts are reported.

Analysis and Results

Table 3 summarizes the results of the analysis. The various factors investigated are identified in the first column. The first figure of the second column shows the number of separate studies reported relating to the factor. The number in parentheses represents the number of observations of effect size derived from the studies. Each treatment group was viewed as a separate response rate observation. Thus, since several studies included multiple treatment groups, the number of observations will be larger than the number of articles. For example, for prenotification by letter, eight articles appeared in the literature, and these eight articles produced 22 observations or measurements of the effect size associated with prenotifying by letter. The third column indicates how many of the individual observations were statistically significant at the 5% risk level (one-sided test). The fourth column is the minimum variance unbiased estimate of effect size. The next two columns show the minimum and maximum effects observed. The next two columns show the standard deviation of the effect estimate of effect size, and the associated z-value for testing the hypothesis that the effect is zero, based on the unbiased minimum variance estimate of effect size. The last column shows the chi-square statistics obtained by combining the individual test significances. As one might expect, the test derived from the overall estimate of effect size, which makes use of the fact that the estimate is normally distributed (asymptotically), appears to be more powerful than the chi-square test reported in the last column, which is basically a "distribution-free" method. However, the chi-square test is more sensitive to individual extreme results. This is the case in stamped versus metered outgoing postage, where one of the experiments yielded a highly significant result causing the chi-square test to be significant while the test based on overall effect was not.

On average, university sponsorship, prenotification by letter, and stamped return postage versus business reply produced the largest increases in response rate. The respective aggregate effect size estimates are 8.9%, 7.7%, and 6.2%. All three effects were significant ($p <$ .01). The aggregate effect associated with postcard follow-up was about 3.5%, also statistically significant ($p < .01$). Very small, but statistically significant ($p < .05$), effects were associated with using

Table 3. Statistical Summary

Factor	Number of Studies[a]	Number of Significant Observations[b]	Effect			Standard Deviation of Effect	z-value	χ^2 (df)
			Combined Estimate	Min.	Max.			
Prenotification by letter	8 (22)	9	.077	−.110	.231	.0099	7.80**	180.30 (44)**
Follow-up by postcard	6 (36)	11	.035	−.110	.240	.0071	4.93**	172.91 (72)**
Outgoing postage								
First class vs. 2d, 3d, or bulk rate (control)	5 (18)	1	.018	−.090	.100	.0075	2.4*	48.77 (36)
Stamped vs. metered (control)	4 (20)	1	.009	−.100	.158	.0090	1.02	56.78 (40)*
Stamped return postage vs. business reply (control)	9 (50)	5	.062	−.100	.320	.0059	10.44**	170.86 (100)**
Notification of cutoff date	5 (12)	1	.017	−.135	.078	.0152	1.12	28.57 (24)
University sponsorship	4 (41)	22	.089	−.112	.225	.0083	10.78**	315.69 (82)**
Color of the questionnaire: green vs. white (control)	3 (10)	0	.020	−.064	.091	.0109	1.83*	28.06 (20)
Postscript asking for cooperation	2 (6)	0	−.024	−.112	.079	.0245	−.98	8.04 (12)

[a] Parenthesized number indicates number of observations on effect.
[b] 5% risk level, one-sided test.
*Statistically significant at 5% risk lelvel ($p < .05$).
**Statistically significant at 1% risk level ($p < .01$).

first-class versus second-class, third-class, or bulk-rate outgoing postage and with using a green, as opposed to white, questionnaire.

Contrary to expectations, stamped outgoing postage did not generate meaningfully higher response rates than did metered postage. The combined statistical evidence is barely significant at the 5% risk level, and the aggregate effect is not statistically significant. This is consistent with the findings of Armstrong and Lusk (1987). The notification of a cutoff date for inclusion of the response did not significantly increase response rate either. Also, a personalized postscript asking for cooperation did not significantly increase response rate. The aggregate effect size was negative, but not statistically significant.

For follow-up by postcard, stamped return postage, and university sponsorship, a substantial number of observations were available, which allowed the individual standardized differences of the observations from the combined estimate of the effect to be examined. In each case, the standardized differences resembled a random sample of independent standard normal random variables indicating that substantial and different interactions were not present in the measurements of the effect. Assuming additivity and no interactions, the impact of using several response stimulating techniques in concert can be quantified. For example, prenotifying respondents by letter and using stamped return postage is estimated to add about 13.9% to the response rate on average (7.7% + 6.2%).

Using weighted least squares to fit the incentive model, $y = \beta\sqrt{x} + \epsilon$, yields an estimate of .031 for the parameter β, and a corresponding statistically significant t-value of 15.2. (The usual R^2 is not reported because it is not a meaningful indicator of fit when the model does not include an intercept.) Hence, an incentive of 25¢ roughly increases response rate by about 16% on average, while an incentive of $1 roughly adds about 31% on average to the response rate, and so on. Figure 1 shows a plot of the observed effects versus incentive size and the resulting model fit via regression analysis. The model tends to underestimate for small incentives and overestimate for large incentives. A slightly better fit would be realized using a power function other than the simple square root, for example, .25 as opposed to .5 as the power exponent. However, given that observations are available for only four incentive values, any model is questionable, and it was felt that a marginal improvement in the fit was not worth pursuing.

Summary and Conclusions

The estimated effect sizes for six of nine factors examined were statistically significant. University sponsorship, prenotification by letter,

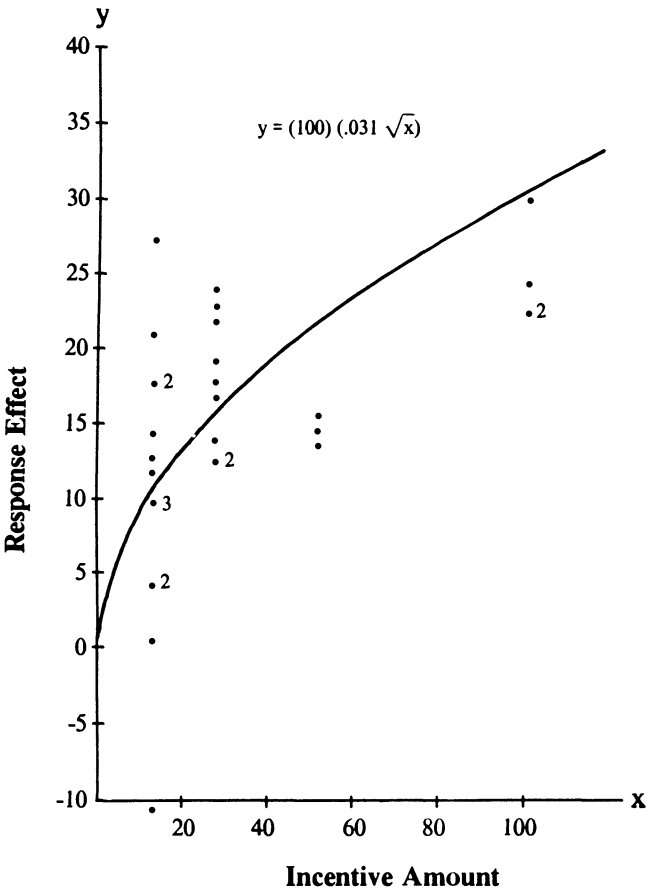

Figure 1. Plot of Estimated Versus Actual Response Effects

stamped return postage, postcard follow-up, first-class outgoing postage, and questionnaire color all successfully increase response rates on average. The most effective factor, sponsorship by a university instead of a private business, is not really controllable, but the other factors are. Notification of a cutoff date and including a postscript asking for cooperation did not significantly increase response rate. While the estimated effect size associated with stamped versus metered outgoing postage was not significant, the accumulated significance (chi-square test) of previous findings was statistically significant.

Including a small cash incentive with the questionnaire was found to have a positive effect on response rate on average. However, the systematic relationship between the marginal increase in response rate

and the size of the incentive indicates that diminishing returns are quickly experienced.

There are differences between the findings of this study and those obtained by the regression approach. For example, Heberlein and Baumgartner (1978) found little or no effect associated with prenotification, while we found that prenotifying respondents of an upcoming survey by postcard increased response rate by almost 8%. Also, differences were observed with respect to the impact of the type of postage used. These differences are likely due to the reasons already cited.

One area needing additional research is interactive or synergistic effects of factors (Wolf, 1986). The minimal evidence gleaned from this meta-analysis suggests that interactions, if they exist, are small. However, we concur with Heberlein and Baumgartner (1978) that a controlled large-scale factorial experiment permitting the estimation of interactions would be a valuable contribution.

References

Armstrong, J. S. (1975)
"Monetary incentives in mail surveys." Public Opinion Quarterly 39:111–116.
Armstrong, J. S., and Edward J. Lusk (1987)
"Return postage in mail surveys." Public Opinion Quarterly 51:233–248.
Churchill, Gilbert A., Jr., Neil M. Ford, Steven W. Hartley, and Orville C. Walker, Jr. (1985)
"The determinants of salesperson performance: A meta-analysis." Journal of Marketing Research 22:103–118.
Dillman, Don A. (1978)
Mail and Telephone Surveys: The Total Design Method. New York: John Wiley and Sons.
Eichner, Klaus, and Werner Habermehl (1981)
"Predicting response rates to mailed questionnaires." American Sociological Review 46:361–363.
Fisher, R. A. (1950)
Statistical Methods for Research Workers. London: Oliver and Boyd.
Glass, Gene V., Barry McGaw, and Mary Lee Smith (1981)
Meta-Analysis in Social Research. Beverly Hills, CA: Sage.
Goyder, John C. (1982)
"Further evidence on factors affecting response rates to mailed questionnaires." American Sociological Review 47:550–553.
Heberlein, Thomas A., and Robert Baumgartner (1978)
"Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature." American Sociological Review 43:447–462.
——— (1981)
"The effectiveness of the Heberlein-Baumgartner models for predicting response rates to mailed questionnaires: European and U.S. examples [reply to Eichner and Habermehl]." American Sociological Review 46:363–367.
Kanuk, Leslie, and Conrad Berenson (1975)
"Mail surveys and response rates: A review." Journal of Marketing Research 12:440–453.
Lentner, Marvin, and Thomas Bishop (1986)
Experimental Design and Analysis. Blacksburg, VA: Valley Book Company.

Linsky, Arnold S. (1975)
"Stimulating responses to mailed questionnaires: A review." Public Opinion Quarterly 39:82–101.
Mazzuca, S. (1982)
"Does patient education for chronic disease have therapeutic value?" Journal of Chronic Diseases 35:521–529.
Rosenthal, Robert (1984)
Meta-Analytic Procedures for Social Research. Applied Social Research Methods, vol. 6. Beverly Hills, CA: Sage.
Wolf, Frederic M. (1986)
Meta-Analysis Quantitative Methods for Research Synthesis. Beverly Hills, CA: Sage.
Yu, Julie, and Harris Cooper (1983)
"A quantitative review of research design effects on response rates to questionnaires." Journal of Marketing Research 20:36–44.

Appendix

Allen, C. T., C. D. Schewe, and G. Wijk (1980)
"More on self-perception theory's foot technique in the pre-call/mail survey setting." Journal of Marketing Research 17:498–502.
Andreasen, A. R. (1970)
"Personalizing mail questionnaire correspondence." Public Opinion Quarterly 34:273–277.
Berdie, D. R. (1973)
"Questionnaire length and response rate." Journal of Applied Psychology 58:278–280.
Blair, E., and G. Ganesh (1984)
"Commitment requests and response rates in three mail surveys." Pp. 368–375 in R. W. Belk, R. Peterson, G. S. Albaum, M. B. Holbrook, R. A. Kerin, N. K. Malhotra, and P. Wright (eds.), 1984 AMA Educators' Proceedings. Chicago: American Marketing Association.
Blumenfeld, W. S. (1973)
"Effect of appearance of corresponding on response rate to a mail questionnaire survey." Psychological Reports 32:178.
*Brook, L. L. (1978)
"The effect of different postage combinations on response levels and speed of reply." Journal of the Market Research Society 20:238–244.
Carpenter, E. H. (1975)
"Personalizing mail surveys: A replication and reassessment." Public Opinion Quarterly 38:614–620.
Childers, T. L., and O. C. Ferrell (1979)
"Response rates and perceived questionnaire length in mail surveys." Journal of Marketing Research 16:429–431.
*Childers, T. L., W. M. Pride, and O. C. Ferrell (1980)
"A reassessment of the effects of appeals on response to mail surveys." Journal of Marketing Research 17:365–370.
Childers, T. L., and S. J. Skinner (1985)
"Theoretical and empirical issues in the identification of survey respondents." Journal of the Market Research Society 27:39–53.
*Cox, E. P., III, W. T. Anderson, Jr., and D. G. Fulcher (1974)
"Reappraising mail survey response rates." Journal of Marketing Research 11:413–417.
*Dillman, D. A. (1972)
"Increasing mail questionnaire response in large samples of the general public." Public Opinion Quarterly 36:254–257.

Dillman, D. A., and J. H. Frey (1974)
"Contribution of personalization to mail questionnaire response as an element of a previously tested method." Journal of Applied Psychology 59:297–301.
Dommeyer, C. J. (1985)
"Does response to an offer of mail survey results interact with questionnaire interest?" Journal of the Market Research Society 27:27–38.
Erdos, P. L., and J. Regier (1977)
"Visible vs. disguised keying on questionnaires." Journal of Advertising Research 17(February):13–18.
Etzel, M. J., and B. J. Walker (1974)
"Effects of alternative follow-up procedures on mail survey response rates." Journal of Applied Psychology 59:219–221.
Feild, H. S. (1975)
"Effects of sex of investigator on mail survey response rates and response bias." Journal of Applied Psychology 60:772–773.
Ferrell, O. C., T. L. Childers, and R. W. Reukert (1984)
"Effects of situational factors on mail survey response." Pp. 364–367 in R. W. Belk, R. Peterson, G. S. Albaum, M. B. Holbrook, R. A. Kerin, N. K. Malhotra, and P. Wright (eds.), 1984 AMA Educators' Proceedings. Chicago: American Marketing Association.
Ford, N. M. (1967)
"The advance letter in mail surveys." Journal of Marketing Research 4:202–204.
*——— (1968)
"Questionnaire appearance and response rates in mail surveys." Journal of Advertising Research 8(September):43–45.
*Friedman, H. M., and A. J. S. Augustine (1979)
"The effects of a monetary incentive and the ethnicity of the sponsor's signature on the rate and quality of response to a mail survey." Journal of the Academy of Marketing Science 7:95–101.
Friedman, H. M., and L. Goldstein (1975)
"Effect of ethnicity of signature on the rate of return and content of a mail questionnaire." Journal of Applied Psychology 60:770–771.
Fuller, C. (1974)
"Effect of anonymity on return rate and response bias in a mail survey." Journal of Applied Psychology 59:292–296.
*Furse, D. H., and D. W. Stewart (1982)
"Monetary incentives versus promised contribution to charity: New evidence on mail survey response." Journal of Marketing Research 19:375–380.
*Furse, D. H., D. W. Stewart, and D. L. Rados (1981)
"Effects of foot-in-the-door, cash incentives, and followups on survey response." Journal of Marketing Research 18:473–478.
*Futrell, C., and R. T. Hise (1982)
"The effects of anonymity and a same-day deadline on the response rate to mail surveys." European Research 10:171–175.
Futrell, C., and J. E. Swan (1977)
"Anonymity and response by salespeople to a mail questionnaire." Journal of Marketing Research 14:611–616.
*Goodstadt, M. S., L. Chung, R. Kronitz, and J. Cook (1977)
"Mail survey response rates: Their manipulation and impact." Journal of Marketing Research 14:391–395.
Goulet, W. M. (1977)
"Efficacy of a third request letter in mail surveys of professionals." Journal of Marketing Research 14:112–114.
*Gullahorn, J. E., and J. T. Gullahorn (1963)
"An investigation of the effects of three factors on response to mail questionnaires." Public Opinion Quarterly 27:294–296.
*Hackler, J. C., and P. Bourgette (1973)
"Dollars, dissonance, and survey returns." Public Opinion Quarterly 37:276–281.

*Hansen, R. A. (1980)
"A self-perception interpretation of the effect of monetary and nonmonetary incentives on mail survey respondent behavior." Journal of Marketing Research 17:77–83.
Hansen, R. A., and L. M. Robinson (1980)
"Testing the effectiveness of alternative foot-in-the-door manipulations." Journal of Marketing Research 17:359–364.
*Harris, J. R., and H. J. Guffey, Jr. (1978)
"Questionnaire returns: Stamps versus business reply envelopes revisited." Journal of Marketing Research 15:290–293.
*Heaton, E. E., Jr. (1965)
"Increasing mail questionnaire returns with a preliminary letter." Journal of Advertising Research 5(December):36–39.
Hendrick C., R. Borden, M. Giesen, E. J. Murray, and B. A. Seyfried (1972)
"Effectiveness of ingratiation tactics in a cover letter on mail questionnaire response." Psychonomic Science 26:349–351.
*Henley, J. R., Jr. (1976)
"Response rate to mail questionnaires with a return deadline." Public Opinion Quarterly 40:374–375.
*Hinrichs, J. R. (1975)
"Factors related to survey response rates." Journal of Applied Psychology 60:249–251.
Hornik, J. (1981)
"Time cue and time perception effect on response to mail surveys." Journal of Marketing Research 18:243–248.
Houston, M. J., and R. W. Jefferson (1975)
"The negative effects of personalization on response patterns in mail surveys." Journal of Marketing Research 12:114–117.
*Houston, M. J., and J. R. Nevin (1977)
"The effects of source and appeal on mail survey response patterns." Journal of Marketing Research 14:374–378.
*Huck, S. W., and E. M. Gleason (1974)
"Using monetary inducements to increase response rates from mailed surveys." Journal of Applied Psychology 59:222–225.
*Jones, W. H., and J. R. Lang (1980)
"Sample composition bias and response bias in a mail survey: A comparison of inducement methods." Journal of Marketing Research 17:69–76.
*Jones, W. H., and G. Linda (1978)
"Multiple criteria effects in a mail survey experiment." Journal of Marketing Research 15:280–284.
Kawash, M. B., and L. M. Aleamoni (1971)
"Effect of personal signature on the initial rate of return of a mailed questionnaire." Journal of Applied Psychology 55:589–592.
*Kernan, J. B. (1971)
"Are 'bulk-rate occupants' really unresponsive?" Public Opinion Quarterly 35:420–424.
*Kimball, A. E. (1961)
"Increasing the rate of return in mail surveys." Journal of Marketing 25(October):63–64.
Labrecque, D. P. (1978)
"A response rate experiment using mail questionnaires." Journal of Marketing 42(October):82–83.
Linsky, A. S. (1965)
"A factorial experiment in inducing responses to a mail questionnaire." Sociology and Social Research 49:183–189.
Mason, W. S., R. J. Dressel, and R. K. Bain (1961)
"An experimental study of factors affecting response to a mail survey of beginning teachers." Public Opinion Quarterly 25:296–299.

*McCrohan, K. F., and L. S. Lowe (1981)
"A cost/benefit approach to postage used on mail questionnaires." Journal of Marketing 45(Winter):130–133.
*McDaniel, S. W., and R. W. Jackson (1984)
"Exploring the probabilistic incentive in mail survey research." Pp. 372–375 in R. W. Belk, R. Peterson, G. S. Albaum, M. B. Holbrook, R. A. Kerin, N. K. Malhotra, and P. Wright (eds.), 1984 AMA Educators' Proceedings. Chicago: American Marketing Association.
McDaniel, S. W., and C. P. Rao (1981)
"An investigation of respondent anonymity's effect on mailed questionnaire response rate and quality." Journal of the Market Research Society 23:150–160.
McGinnis, M. A., and C. J. Hollon (1977)
"Mail survey response rate and bias: The effect of home versus work address." Journal of Marketing Research 14:383–384.
Mizes, J. S., E. L. Fleece, and C. Roos (1984)
"Incentives for increasing return rates: Magnitude levels, response bias, and format." Public Opinion Quarterly 48:794–800.
*Myers, J. H., and A. F. Haug (1969)
"How a preliminary letter affects mail survey returns and costs." Journal of Advertising Research 9(September):37–39.
Nederhof, A. J. (1983)
"The effects of material incentives in mail surveys: Two studies." Public Opinion Quarterly 47:103–111.
*Newman, S. W. (1962)
"Differences between early and late respondents to a mailed survey." Journal of Advertising Research 2(June):37–39.
*Parsons, R. J., and T. S. Medford (1972)
"The effect of advance notice in mail surveys of homogeneous groups." Public Opinion Quarterly 36:258–259.
*Peterson, R. A. (1975)
"An experimental investigation of mail-survey responses." Journal of Business Research 3:199–209.
*Pressley, M. M. (1979)
"Care needed when selecting response inducements in mail surveys of commercial populations." Journal of the Academy of Marketing Science 6:336–343.
*Pressley, M. M., and W. L. Tullar (1977)
"A factor interactive investigation of mail survey response rates from a commerical population." Journal of Marketing Research 14:108–111.
*Pucel, D. J., H. F. Nelson, and D. N. Wheeler (1971)
"Questionnaire follow-up returns as a function of incentives and responder characteristics." Vocational Guidance Quarterly 19:188–193.
*Roberts, R. E., O. F. McCrory, and R. N. Forthofer (1978)
"Further evidence on using a deadline to stimulate responses to a mail survey." Public Opinion Quarterly 42:407–410.
Robertson, D. H., and D. N. Bellenger (1978)
"A new method of increasing mail survey responses: Contributions to charity." Journal of Marketing Research 15:632–633.
*Robin, D. P., and C. G. Walters (1976)
"The effect on return rate of messages explaining monetary incentives in mail questionnaire studies." Journal of Business Communication 13(Spring):49–54.
Roeher, G. A. (1963)
"Effective techniques in increasing response to mailed questionnaires." Public Opinion Quarterly 27:299–302.
Roscoe, A. M., D. Lang, and J. N. Sheth (1975)
"Follow-up methods, questionnaire length, and market differences in mail surveys." Journal of Marketing 39(April):20–27.

Rucker, M., R. Hughes, R. Thompson, A. Harrison, and N. Vanderlip (1984)
"Personalization of mail surveys: Too much of a good thing?" Educational and Psychological Measurement 44:893–905.
Simon, R. (1967)
"Responses to personal and form letters in mail surveys." Journal of Advertising Research 7(March):28–30.
*Stafford, J. E. (1966)
"Influence of preliminary contact on mail returns." Journal of Marketing Research 3:410–411.
Stevens, R. E. (1975)
"Does precoding mail questionnaires affect response rates?" Public Opinion Quarterly 38:621–622.
Tedin, K. L., and C. R. Hofstetter (1982)
"The effect of cost and importance factors on the return rate for single and multiple mailings." Public Opinion Quarterly 46:122–128.
Veiga, J. F. (1974)
"Getting the mail questionnaire returned: Some practical research considerations." Journal of Applied Psychology 59:217–218.
*Vocino, T. (1977)
"Three variables in stimulating responses to mailed questionnaires." Journal of Marketing 41(October):76–77.
*Walker, B. J., and R. K. Burdick (1977)
"Advance correspondence and error in mail surveys." Journal of Marketing Research 14:379–382.
*Watson, J. J. (1965)
"Improving the response rate in mail research." Journal of Advertising Research 5(June):48–50.
Whitmore, W. J. (1976)
"Mail survey premiums and response bias." Journal of Marketing Research 13:46–50.
Wildman, R. C. (1977)
"Effects of anonymity and social setting on survey responses." Public Opinion Quarterly 41:74–79.
*Wiseman, F. (1973)
"Factor interaction effects in mail survey response rates." Journal of Marketing Research 10:330–333.
*Wotruba, T. R. (1966)
"Monetary inducements and mail questionnaire response." Journal of Marketing Research 3:398–400.
Wynn, G. W., and S. W. McDaniel (1985)
"The effect of alternative foot-in-the-door manipulations on mailed questionnaire response rate and quality." Journal of the Market Research Society 27:15–26.