UNDERSTANDING MAIL SURVEY RESPONSE BEHAVIOR
A META-ANALYSIS

FRANCIS J. YAMMARINO
STEVEN J. SKINNER
TERRY L. CHILDERS

Abstract  A meta-analysis of prior studies of techniques designed to induce mail survey response rates was conducted. Research encompassing 184 effects (study outcomes) in 115 studies (articles) for 17 predictors of response rate was examined. The average effect size across all manipulations was \( r = .065 \), indicating an average increase of about 6.5 percent in response rates for manipulations. Effect sizes for specific predictors and two potential moderators of effects were examined. Results indicated that repeated contacts in the form of preliminary notification and follow-ups, appeals, inclusion of a return envelope, postage, and monetary incentives, were effective in increasing survey response rates. Significant effect sizes for the predictors ranged from an increase in response of 2 percent to 31 percent. Implications of the results for the conduct of mail surveys and future research on mail survey response behavior are discussed.

Researchers have amassed myriad techniques to stimulate mail survey response rates, reduce item omission, speed up response, and reduce response bias. Examples include the use of preliminary notification, follow-up, sponsorship, appeals, postage, personalization, incentives, anonymity, prior commitment, and techniques affecting questionnaire

Correspondence regarding this article should be addressed to FRANCIS J. YAMMARINO, School of Management and Center for Leadership Studies, State University of New York at Binghamton, P.O. Box 6000, Binghamton, NY 13902-6000. STEVEN J. SKINNER is at the College of Business and Economics, University of Kentucky, Lexington, KY. TERRY L. CHILDERS is at the Carlson School of Management, University of Minnesota, Minneapolis, MN. This research was supported in part by a New Faculty Development Award to F. J. Yammarino from the NYS/UUP Professional Development and Quality of Working Life Committee. The views expressed are ours. We gratefully acknowledge the helpful comments provided by the anonymous reviewers on an earlier version of this article.

All rights reserved. 0033-362X/91/5504-0004$02.50
appearance. These methods represent manipulations or predictors which, at least in some instances, have served to facilitate survey participation. A comprehensive list and review of studies related to the effectiveness of these methods can be found in the work of Duncan (1979), Houston and Ford (1976), Kanuk and Berenson (1975), and Linsky (1975).

Linsky (1975), for example, found strong evidence that preliminary notification, follow-ups, "high powered" postage, sponsorship, small monetary incentives, and nonmonetary incentives were very effective for, or had positive effects on, increasing response rates. Other manipulations, according to Linsky (1975), are equivocal or interact with additional factors to affect response rates. Likewise, Kanuk and Berenson (1975) determined that monetary incentives (especially small ones), follow-up contacts, and perhaps preliminary notification, were the only techniques that had strong empirical evidence for effectiveness in increasing response rates. Other techniques showed no consistent effect.

In a recent qualitative review, Harvey (1987) found that follow-ups, preliminary notification, stamped reply envelopes, and monetary incentives were important factors for increasing response rates. Harvey (1987) also determined that a number of other techniques such as personalization, length, sponsorship, and appeals had a more ambiguous effect on response rates.

Although useful, traditional qualitative reviews are limited in that they do not explicitly consider differences in studies due to sample size, sample composition, experimental executions, and effect sizes (Fox, Crask, and Kim 1988). Recently, quantitative reviews such as meta-analysis, designed to address these limitations, have emerged. Several of these meta-analyses provide important background for the present study and are summarized in Table 1.

In an early comprehensive review of the mail survey literature, Heberlein and Baumgartner (1978) examined nine factors affecting initial and final survey response rates across 98 studies. They were able to explain 51 percent of the variance in final response through the number of respondent contacts (preliminary and follow-up) and the saliency of the survey topic. Additionally, Heberlein and Baumgartner (1978) found that sponsorship, length of the questionnaire, type of population surveyed, incentives, and postage (metered or franked) were also significant determinants of response (see Table 1). Goyder (1982) replicated the Heberlein and Baumgartner (1978) study by conducting a meta-analysis on 145 citations. His findings were largely consistent with the predictive model of Heberlein and Baumgartner, finding that type of population, number of contacts, sponsorship, and
monetary incentives were most important in predicting final survey response rate.

In another quantitative review of 93 studies for all types of surveys (personal, telephone, and mail), Yu and Cooper (1983) examined 15 factors reported potentially to affect survey participation rates. They found, across method of contact, that response rates were improved by incentives (prepaid and promised), nonmonetary premiums and rewards, and by increasing the amount of monetary rewards. Other techniques that increased response were preliminary notification, foot-in-the-door approaches, personalization, and follow-up contacts (table 1).

In a recent meta-analysis, Fox, Crask, and Kim (1988) examined 82 studies with respect to eight factors that have been associated across studies with higher response rates to mail surveys. In their analysis, prenotification, follow-ups, and stamped return postage were found to be significant determinants of mail survey response rates (table 1). Additionally, sponsorship and small monetary incentives enhanced response.

Two other more focused quantitative reviews examined the effects of personalization (Worthen and Valcarce 1985) and several types of return postage (Armstrong and Lusk 1987). Worthen and Valcarce (1985) reviewed 26 studies and concluded that the effect size for personalization from previous research was small, and in a follow-up experiment found no effects on response rate due to cover letter personalization. Armstrong and Lusk (1987) conducted their meta-analysis on 34 studies and concluded that the only effect of consequence on response rate was first-class postage as compared to business reply, and even that effect was weak (table 1).

In synthesizing across the prior meta-analyses, a subset of the mail survey response determinants appears to be consistently predictive of response rates. Drawing from the above studies, repeated contacts, whether they be in the form of preliminary notification of a survey or a survey follow-up, appear to have a considerable positive impact on response rates (Fox, Crask, and Kim 1988; Goyder 1982; Heberlein and Baumgartner 1978; Yu and Cooper 1983). Another factor that also has been reported to have a consistent positive effect is the use of monetary incentives (Fox, Crask, and Kim 1988; Goyder 1982; Heberlein and Baumgartner 1978; Yu and Cooper 1983). In addition, the sponsor of a survey, although less under the control of the researcher, has been a key determinant of survey response (Fox, Crask, and Kim 1988; Goyder 1982; Heberlein and Baumgartner 1978). Finally, the inclusion of and type of outgoing and return postage has been observed consistently to affect response rates (Armstrong and Lusk 1987; Fox, Crask, and Kim 1988).
Table 1. Summary of Meta-analyses on Mail Survey Response Rates

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of studies</td>
<td>115</td>
<td>98</td>
<td>93</td>
<td>34</td>
<td>82</td>
</tr>
<tr>
<td>No. of effects</td>
<td>184</td>
<td>214</td>
<td>497</td>
<td>34</td>
<td>214</td>
</tr>
<tr>
<td>Publication years</td>
<td>1940–88</td>
<td>NR(^a)</td>
<td>1965–81</td>
<td>1951–86</td>
<td>1961–86</td>
</tr>
<tr>
<td>Predictors:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preliminary notification</td>
<td>X*</td>
<td>X*</td>
<td>X*</td>
<td>...</td>
<td>X*</td>
</tr>
<tr>
<td>Follow-ups/repeated contacts</td>
<td>X*</td>
<td>X*</td>
<td>X*</td>
<td>...</td>
<td>X*</td>
</tr>
<tr>
<td>Sponsorship</td>
<td>X</td>
<td>X*</td>
<td>X(^b)</td>
<td>...</td>
<td>X*</td>
</tr>
<tr>
<td>Appeals</td>
<td>X*</td>
<td>...</td>
<td>X</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Reply/return envelope</td>
<td>X*</td>
<td>...</td>
<td>X(^b)</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Postage</td>
<td>X*</td>
<td>X*</td>
<td>X</td>
<td>X*</td>
<td>X*</td>
</tr>
<tr>
<td>Personalization</td>
<td>X</td>
<td>X</td>
<td>X*</td>
<td>...</td>
<td>X</td>
</tr>
<tr>
<td>Monetary incentives</td>
<td>X*</td>
<td>X*</td>
<td>X*</td>
<td>...</td>
<td>X*</td>
</tr>
<tr>
<td>Feature</td>
<td>X</td>
<td>...</td>
<td>X*</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Nonmonetary incentives</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survey appearance</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anonymity/no identification</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>X*</td>
<td>X*</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deadlines</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic of survey</td>
<td>...</td>
<td>X*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of sample</td>
<td>X*</td>
<td>X*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Offering survey results</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foot in the door</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td>X*</td>
</tr>
<tr>
<td>Cover letter</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Questionnaire color</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td>X*</td>
</tr>
</tbody>
</table>

Note.—The Goyder (1982) study was not included in this table because it essentially replicated the Heberlein and Baumgartner (1978) study, and the results as well as factors studied basically mirrored their work. The Worthen and Valcarce (1985) study also was not included in this table because it did not employ traditional meta-analysis techniques; rather, a sign test on the direction of effects was used in their work. "X" indicates factor was examined in study.

a Information not reported in the article.
b Effect could not be tested in this article.
* Statistically significant effect.
The present study complements and extends previous work, including the meta-analyses of Armstrong and Lusk (1987), Fox, Crask, and Kim (1988), Goyder (1982), Heberlein and Baumgartner (1978), Worthen and Valcarce (1985), and Yu and Cooper (1983), in several ways. First, a meta-analysis is used to assess more factors (17; see table 2) which influence response rate than were examined by previous researchers. As summarized in table 1, the present research found several of these factors to have a significant effect on survey response (see details below). Second, a larger number of study outcomes (184 effects in 115 articles) than in most prior research are included in the current meta-analysis. Third, different from previous work, two moderator variables (year of publication and type of sample) which may have affected the results of previous studies are examined. Given previous work (Harvey 1987; Heberlein and Baumgartner 1978; Yu and Cooper 1983), the publication year of the study was included as a type of cohort effect because of the changes in respondents’ habits and journal editors’ preferences over time, and to reflect any potential accumulation of survey administration knowledge. The type of subject sampled (e.g., consumer, educational, industrial) was included as a potential influence on response rates given work and comments on subjects and target populations (Goyder 1982; Harvey 1987; Heberlein and Baumgartner 1978; Kanuk and Berenson 1975; Linsky 1975). In these ways, the purpose of this study was to enhance understanding of mail survey response behavior by presenting a more comprehensive meta-analysis investigation of factors that affect mail survey response rates.

**Meta-analysis Method**

To accomplish this purpose, a quantitative review of research design effects on mail survey response rates was conducted using meta-analysis procedures described by Glass (1977), Hunter and Schmidt (1990), and Hunter, Schmidt, and Jackson (1982). The procedures were used to determine mean effect sizes for factors which affect response rate and whether these effects are the same for different target audiences and whether the effects have changed over time. Also, based on these procedures, a determination can be made whether variation in results across samples (studies) was due to methodological problems and statistical artifacts (e.g., sampling error) as compared to actual substantive differences in subpopulation correlations of response rates and design effects identified by underlying moderator variables (e.g., target groups or time periods) (see details below). Response rate, the dependent variable, was defined as response as a percent of size of
contacted sample (Yu and Cooper 1983). The independent variables were the manipulated factors designed to affect response rate.

**STUDY INCLUSION AND CODING**

Individual studies (articles) that were included in the meta-analysis were identified in several ways. First, reference lists of several previously published review articles on mail surveys were examined (e.g., Fox, Crask, and Kim 1988; Harvey 1987; Heberlein and Baumgartner 1978; Kanuk and Berenson 1975; Linsky 1975; Yu and Cooper 1983). Second, a multi-abstracting services computer research was conducted for all available years in the data bases. These included ABI/Inform (1971+), ERIC (1966+), Social SciSearch (1972+), Sociological Abstracts (1963+), and PsInfo (1967+). Key words used in the search included “mail surveys,” “response rates,” and each of these terms combined with the names and synonyms for the various factors listed in table 1. Third, for recent work (1978+), a manual search of 10 journals that have a history of publishing mail survey studies was performed (i.e., American Sociological Review, Journal of the Academy of Marketing Science, Journal of Advertising Research, Journal of the American Statistical Association, Journal of Applied Psychology, Journal of Marketing, Journal of Marketing Research, Journal of the Market Research Society, Psychological Reports, and Public Opinion Quarterly). The manual search uncovered three articles not identified from previously published reference lists or from the computer search.

To be included in the meta-analysis, studies must have manipulated a factor that influences response rates and reported response rates for the different experimental conditions. If more than one factor was manipulated in a study, each individual effect (study outcome) was included as a separate “data point” for the meta-analysis. Over 25 journals from disciplines such as business, education, marketing, political science, psychology, sociology, and statistics were represented. The 10 journals that yielded the most articles for the meta-analysis were: Public Opinion Quarterly, 27 articles; Journal of Marketing Research, 18 articles; Journal of Applied Psychology, 11 articles; Journal of Marketing and Journal of the Market Research Society, 8 articles each; Journal of Advertising Research, 7 articles; American Sociological Review and Journal of the Academy of Marketing Science, 4 articles each; Journal of the American Statistical Association and Psychological Reports, 3 articles each. The remaining journals produced 1 or 2 articles for the meta-analysis. The time period was from 1940 to January 1988, and 115 articles yielding 184 data points (effects) for 17 different predictors of response rate were included. Although many
articles yielded multiple data points, the effects were always for different predictors of response rate. Thus, for any given predictor, only one data point per study was included in the meta-analysis. The articles included in the meta-analysis are listed in the Appendix, and the 17 manipulated factors of interest are identified in table 2.

All information relevant to the meta-analysis was coded from the articles by four trained coders. Data from each article were coded by two individuals independently. The coders then compared their information. In cases of disagreement, a third coder was asked to code the article in question independently and then all three coders met to determine a consensus coding for the article. Although inter-rater reliability statistics were not calculated, there were very few instances of nonagreement between raters.

Response rate, the factor that was manipulated (in general, a "yes/no" code was used), sample size, statistics, effect sizes, and direction of relationships were coded. The two moderator variables also were coded. To ensure an adequate number of data points for meta-analysis for the predictors of response rates under several conditions of the moderators, publication years were grouped as 1940–70, 1971–75, and 1976–87, while sample type was classified to contrast consumers versus various institutional groups (see details below).

META-ANALYSIS CALCULATIONS

To distinguish between artifactual (e.g., sampling error) and substantive sources of variation, the procedures developed by Glass (1977), Hunter and Schmidt (1990), and Hunter, Schmidt, and Jackson (1982) were combined in a computer program specifically designed for the meta-analysis in this study. The work of Glass (1977, p. 374) was used to convert study statistics (e.g., t, F, $X^2$, contingency table, Mann-Whitney U) to uncorrected product-moment correlations. Measurement error (reliability) and range restriction affecting these correlations were generally not issues of concern in the present study because the variables involved were actual counts of mailings and responses. Except for potential coding and typographical errors, therefore, reliability should be 1.00 and the range should be completely unrestricted. The frequency-weighted mean correlation ($r$) across studies, the best estimate of the population correlation ($p$), and the frequency-weighted average squared error ($S^2$), the corresponding uncorrected or observed variance of correlations across studies, were then computed using the work of Hunter and Schmidt (1990) and Hunter, Schmidt, and Jackson (1982). Because the observed or uncorrected total variance ($S^2$) can be confounded by variation in sample correlations produced by sampling error, it was corrected by removing sampling error.
variance as described by Hunter and Schmidt (1990) and Hunter, Schmidt, and Jackson (1982) to derive the best estimate of the population variance ($\sigma^2_p$).

To test the significance of the variance corrected for sampling error or unexplained variance, a $\chi^2$ statistic was used (Hunter and Schmidt 1990; Hunter, Schmidt, and Jackson 1982). Because of very high statistical power, however, even very small amounts of variation across studies will be statistically significant. That is, variation may be of minimal magnitude and significant, but when the $\chi^2$ is nonsignificant, strong support is provided for the lack of actual variation across studies. As suggested by Hunter and Schmidt (1990) and Hunter, Schmidt, and Jackson (1982), when the unexplained (corrected) variance is small, inconsistent results across studies are due mainly to statistical artifacts, and subpopulation correlations do not differ. Thus, it is not necessary to research for moderator variables. In contrast, when the corrected (unexplained) variance is large, inconsistent results across studies are not artifactual, and they recommend examination of moderator variables (year of publication and type of sample in the present study) to identify subpopulations with different correlational values.

As a general rule, a moderator search should be initiated if the ratio of unexplained (corrected) to total variance is greater than 25 percent, or in contrast, if the ratio of explained (sampling error) to total variance is less than 75 percent. Regardless of other indicators (e.g., statistical significance of mean correlations), this 75 percent rule is used to determine whether a moderator search is warranted (Hunter and Schmidt 1990; Hunter, Schmidt, and Jackson 1982). The above calculations and procedures are then repeated within each condition (level) of the potential moderator variables. If (1) the frequency-weighted mean correlation varies from subgroup to subgroup and (2) the average corrected variance for the subgroups is lower than the corrected variance for the data set as a whole, then there is evidence for a moderated effect. Clearly, after examining one moderator, the unexplained (corrected) variance may still be greater than zero for each level of a moderator. This suggests that the remaining variance is due to additional moderators or statistical artifacts.

Results

OVERALL RESULTS

The findings of meta-analysis for each of the research design effects (predictors) on mail survey response rates are presented in table 2.
<table>
<thead>
<tr>
<th>Predictor</th>
<th>Correlation(^a)</th>
<th>95% C.I.</th>
<th>Corrected Variance</th>
<th>(X^2)</th>
<th>No. of Correlations</th>
<th>Range of Evaluation</th>
<th>Total Sample Size</th>
<th>Uncorrected Variance</th>
<th>Variance Explained(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>.065</td>
<td>.106 to</td>
<td>.236</td>
<td>.008</td>
<td>2467.59***</td>
<td>.217 to .850</td>
<td>296.601</td>
<td>.008</td>
<td>.075</td>
</tr>
<tr>
<td>Preliminary notification</td>
<td>.176</td>
<td>.122 to</td>
<td>.474</td>
<td>.023</td>
<td>244.27***</td>
<td>.106 to .446</td>
<td>9.265</td>
<td>.025</td>
<td>.066</td>
</tr>
<tr>
<td>Follow-ups/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>repeated contacts</td>
<td>.156</td>
<td>-.083 to</td>
<td>.395</td>
<td>.015</td>
<td>307.88***</td>
<td>.014 to .319</td>
<td>19.055</td>
<td>.015</td>
<td>.032</td>
</tr>
<tr>
<td>Sponsorship</td>
<td>.037</td>
<td>-.049 to</td>
<td>.123</td>
<td>.002</td>
<td>64.52***</td>
<td>.016 to .145</td>
<td>30.188</td>
<td>.002</td>
<td>.093</td>
</tr>
<tr>
<td>Appeals</td>
<td>.047</td>
<td>.010 to</td>
<td>.083</td>
<td>.000</td>
<td>3.81</td>
<td>.028 to .074</td>
<td>2.321</td>
<td>.002</td>
<td>.787</td>
</tr>
<tr>
<td>Reply/return envelope</td>
<td>.079</td>
<td>.005 to</td>
<td>.154</td>
<td>.001</td>
<td>49.51***</td>
<td>.029 to .222</td>
<td>22.664</td>
<td>.002</td>
<td>.323</td>
</tr>
<tr>
<td>Return postage:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stamped/metered</td>
<td>.022</td>
<td>-.031 to</td>
<td>.074</td>
<td>.001</td>
<td>36.62**</td>
<td>.051 to .117</td>
<td>22.967</td>
<td>.002</td>
<td>.546</td>
</tr>
<tr>
<td>Special delivery/air mail</td>
<td>.028</td>
<td>-.012 to</td>
<td>.069</td>
<td>.000</td>
<td>12.18</td>
<td>.029 to .117</td>
<td>7.430</td>
<td>.002</td>
<td>.739</td>
</tr>
<tr>
<td>Personalization</td>
<td>.037</td>
<td>-.070 to</td>
<td>.143</td>
<td>.003</td>
<td>145.15***</td>
<td>-.197 to .244</td>
<td>37.465</td>
<td>.004</td>
<td>.234</td>
</tr>
<tr>
<td>Incentives:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\leq .50)</td>
<td>.184</td>
<td>.094 to</td>
<td>.274</td>
<td>.002</td>
<td>96.93***</td>
<td>.028 to .263</td>
<td>36.217</td>
<td>.002</td>
<td>.155</td>
</tr>
<tr>
<td>&gt; .50 and (\leq 1.00)</td>
<td>.119</td>
<td>-.138 to</td>
<td>.376</td>
<td>.017</td>
<td>24.65***</td>
<td>.025 to .434</td>
<td>1.168</td>
<td>.021</td>
<td>.162</td>
</tr>
<tr>
<td>&gt; 1.00</td>
<td>.122</td>
<td>.070 to</td>
<td>.175</td>
<td>.001</td>
<td>8.91</td>
<td>-.049 to .173</td>
<td>3.892</td>
<td>.002</td>
<td>.673</td>
</tr>
<tr>
<td>Nonmonetary</td>
<td>.075</td>
<td>-.020 to</td>
<td>.169</td>
<td>.002</td>
<td>33.11***</td>
<td>-.075 to .235</td>
<td>9.012</td>
<td>.004</td>
<td>.362</td>
</tr>
<tr>
<td>Survey appearance</td>
<td>.017</td>
<td>-.026 to</td>
<td>.060</td>
<td>.000</td>
<td>13.96*</td>
<td>-.023 to .152</td>
<td>14.421</td>
<td>.001</td>
<td>.501</td>
</tr>
<tr>
<td>Anonymity/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no identification</td>
<td>.021</td>
<td>-.089 to</td>
<td>.131</td>
<td>.003</td>
<td>246.12***</td>
<td>-.047 to .850</td>
<td>74.259</td>
<td>.003</td>
<td>.045</td>
</tr>
<tr>
<td>Lengths:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\leq 4) pages</td>
<td>.010</td>
<td>-.267 to</td>
<td>.288</td>
<td>.020</td>
<td>20.23***</td>
<td>-.217 to .255</td>
<td>809</td>
<td>.025</td>
<td>.198</td>
</tr>
<tr>
<td>&gt; 4 pages</td>
<td>-.078</td>
<td>-.078 to</td>
<td>-.078</td>
<td>.000</td>
<td>5.85</td>
<td>-.122 to -.015</td>
<td>3.421</td>
<td>.002</td>
<td>1.000</td>
</tr>
<tr>
<td>Deadlines</td>
<td>.027</td>
<td>-.044 to</td>
<td>.098</td>
<td>.001</td>
<td>5.69</td>
<td>-.061 to .066</td>
<td>2.047</td>
<td>.003</td>
<td>.527</td>
</tr>
</tbody>
</table>

Note.—Correlations and variances rounded to three decimal places.
\(^a\) Frequency-weighted mean correlation.
\(^b\) Variance-weighted mean correlation.
* \(p < .05\).
** \(p < .01\).
*** \(p < .001\).
Table 3. Distribution of 184 Correlations for Meta-analysis of Mail Survey Response Rates

<table>
<thead>
<tr>
<th>Range of Correlations</th>
<th>Number of Correlations</th>
<th>Percent of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; -.201</td>
<td>1</td>
<td>.5</td>
</tr>
<tr>
<td>-.200 to -.101</td>
<td>5</td>
<td>2.7</td>
</tr>
<tr>
<td>-.100 to -.001</td>
<td>32</td>
<td>17.4</td>
</tr>
<tr>
<td>.000 to .100</td>
<td>84</td>
<td>45.7</td>
</tr>
<tr>
<td>.101 to .200</td>
<td>40</td>
<td>21.7</td>
</tr>
<tr>
<td>.201 to .300</td>
<td>12</td>
<td>6.5</td>
</tr>
<tr>
<td>.301 to .400</td>
<td>7</td>
<td>3.8</td>
</tr>
<tr>
<td>.401 to .500</td>
<td>2</td>
<td>1.1</td>
</tr>
<tr>
<td>.501 to .600</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>.601 to .700</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>.701 to .800</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>&gt; .800</td>
<td>1</td>
<td>.5</td>
</tr>
</tbody>
</table>

For each predictor shown in the first column of the table, the following information is presented in columns 2–9: frequency-weighted mean correlation; 95 percent confidence interval for the correlation; corrected (unexplained) variance (i.e., total minus sampling error variance); \( \chi^2 \) test of corrected (unexplained) variance; number of correlations (effects); range of the correlations (effects); total sample size; uncorrected total or observed variance in the sample correlations; and variance explained by sampling error (i.e., ratio of sampling error to total variance).

Results collapsed across all design effects are shown in the first row of table 2; that is, regardless of the manipulation in a study, the effect (correlation) was .065, which was not statistically significant (95 percent confidence interval includes zero). The range of the correlations in this case was -.217 to .850. The 184 correlations were distributed as shown in table 3. With the exception of one outlier, the correlations were generally well dispersed throughout the range and somewhat normally distributed. As indicated by the remaining data in the first row of table 2, subpopulation correlations should be examined, and the logical starting point is with the 17 different manipulations also identified in table 2.

Inspection of the distributions of the correlations for each of the 17 manipulations shown in table 2 revealed two somewhat interesting results. First, for incentives greater than $..50 and less than or equal to $1.00, the frequency-weighted mean correlation of .119 was ob-
tained from two significant effects (.434 and .308) and two nonsignificant effects (.025 and .043). This "split" distribution and small number of correlations may thus indicate an unstable finding. Second, for anonymity, the frequency-weighted mean correlation of .021 was obtained from 10 effects within the −.047 to .089 range and one outlier effect (.850). But this outlier effect was based on a small number of subjects (300 out of 74,259), so the overall effect size was not inflated artificially. For the remaining manipulations, the effects were fairly well distributed across each of the ranges reported in table 2. Of the 17 correlations, only the correlations for appeals (.047), reply/return envelope (.079), incentives of $.50 or less (.184), incentives greater than $1.00 (.122), and survey length greater than four pages (−.078) differed significantly from zero (95 percent C.I.).

For appeals and survey length greater than four pages, χ² tests of corrected (unexplained) variance were not significant, and more than 75 percent of total (uncorrected) variance was explained by sampling error variance. A moderator search for these two predictors of response rate was not necessary—the correlations for appeals and survey length greater than four pages were significant and derived from one population. Although χ² tests for special delivery/air mail return postage, incentives greater than $1.00, and deadlines were not significant, the ratio of sampling error to total variance for each is less than 75 percent. For the remaining manipulations identified in table 2, χ² tests were significant, and less than 75 percent of total variance was explained by sampling error variance. Therefore, except for appeals and survey length greater than four pages, the sample correlations were derived from more than one population, and moderators need to be investigated.

MODERATOR RESULTS

Subpopulation correlations were investigated for two potential moderators—year of publication and type of sample—of the relationship between research design effects and mail survey response rates. Frequencies of correlations across all predictors of response rate for different levels of the two moderators were examined. The final categorization of year of publication was 1940–70 (62 correlations), 1971–75 (55 correlations), and 1976–87 (67 correlations). Type of sample was originally coded as consumer (108 correlations), educational (35 correlations), industrial (19 correlations), health care (10 correlations), governmental (4 correlations), and other institutional (8 correlations) groups. The final categorization of type of sample was consumer and institutional groups (108 and 76 correlations, respectively). These final categorizations were chosen to ensure an adequate number of correla-
tions per cell for meta-analysis of individual predictors as well as to identify two independent moderators. Nonsignificant $\chi^2$ results (2.116) indicated that the frequencies of correlations did not differ by cell, and the two moderator coding schemes, year of publication and type of sample, were independent of one another ($r = .093, \text{n.s.}; \phi = .107$). Thus, two separate ways of specifying subpopulations were identified.

Moderator analyses were conducted for all predictors of response rates for which variance explained by sampling error was less than 75 percent and which had sufficient numbers of correlations (three or more) in cells based on time periods or types of samples to perform meta-analysis. To conserve space, only results of comparisons using the previously described criteria for determining a significant moderator are presented in table 4. The same type of information about correlations and variances as displayed in table 2 is shown in table 4 for predictors of mail survey response rates by year of publication and by type of sample moderator.

For four predictors of response rate (preliminary notification, reply/return envelope, stamped/metered postage, and personalization), all three levels of the year of publication moderator were represented. In only two cases, however, did there seem to be a moderated effect. For preliminary notification, the subgroup correlations varied from .010 (1971-75) to .285 (1976-87), the latter one was significantly different from zero, and the average corrected variance in the subgroups was .014 as compared to a corrected variance of .023 for the data as a whole. Although variance explained by sampling error for the subgroups did not reach 75 percent (suggesting the presence of additional moderators), year of publication did appear to moderate the preliminary notification-response rate relationship. For stamped/metered return postage, the subgroup correlations ranged from -.008 (1971-75) to .066 (1940-70), all three differed significantly from zero, the average corrected variance in the subgroups was zero as compared to a corrected variance of .001 for the data as a whole, and variance explained by sampling error was 100 percent in all subgroups. Year of publication, therefore, did appear to be a significant moderator of the stamped/metered return postage-response rate relationship.

For eight design effects on response rates (preliminary notification, follow-ups/repeated contacts, stamped/metered postage, personalization, incentives less than or equal to $.50, nonmonetary incentives, survey appearance, and anonymity), both levels of the type of sample moderator were represented. In only two cases, however, did there seem to be a moderated effect. For follow-ups/repeated contacts, the subgroup correlations were .073 (consumer groups) and .306 (institutional groups), the latter one was significantly different from zero, and the average corrected variance in the subgroups was .002 as compared
Table 4. Meta-analysis Results for Moderators of Mail Survey Response Rates

<table>
<thead>
<tr>
<th>Predictor and Moderator</th>
<th>Correlation*</th>
<th>95% C.I.</th>
<th>Corrected Variance</th>
<th>$X^2$</th>
<th>No. of Correlations</th>
<th>Range of Correlations</th>
<th>Total Sample Size</th>
<th>Uncorrected Variance</th>
<th>Variance Explainedb</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year of publication:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preliminary notification:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1940–70</td>
<td>.165</td>
<td>-.119 to .448</td>
<td>.021</td>
<td>133.01**</td>
<td>8</td>
<td>.028 to .466</td>
<td>5.658</td>
<td>.022</td>
<td>.060</td>
</tr>
<tr>
<td>1971–75</td>
<td>.010</td>
<td>-.217 to .237</td>
<td>.013</td>
<td>20.18**</td>
<td>4</td>
<td>-.106 to .214</td>
<td>1.206</td>
<td>.017</td>
<td>.198</td>
</tr>
<tr>
<td>Stamped/metered return postage:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1940–70</td>
<td>.066</td>
<td>.066 to .066</td>
<td>.000</td>
<td>1.89</td>
<td>4</td>
<td>.032 to .083</td>
<td>6.257</td>
<td>.000</td>
<td>1.000</td>
</tr>
<tr>
<td>1971–75</td>
<td>-.008</td>
<td>-.008 to -.008</td>
<td>.000</td>
<td>3.86</td>
<td>5</td>
<td>-.044 to .026</td>
<td>10.123</td>
<td>.000</td>
<td>1.000</td>
</tr>
<tr>
<td>1976–87</td>
<td>.024</td>
<td>.024 to .024</td>
<td>.000</td>
<td>10.04</td>
<td>11</td>
<td>-.051 to .117</td>
<td>6.587</td>
<td>.002</td>
<td>1.000</td>
</tr>
<tr>
<td><strong>Type of sample:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follow-ups/ repeated contacts:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consumer groups</td>
<td>.073</td>
<td>-.043 to .189</td>
<td>.003</td>
<td>50.24**</td>
<td>7</td>
<td>.014 to .316</td>
<td>12.256</td>
<td>.004</td>
<td>.139</td>
</tr>
<tr>
<td>Institutional groups</td>
<td>.306</td>
<td>.272 to .341</td>
<td>.000</td>
<td>5.60</td>
<td>3</td>
<td>.251 to .319</td>
<td>6.799</td>
<td>.001</td>
<td>.536</td>
</tr>
<tr>
<td>Stamped/metered return postage:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consumer groups</td>
<td>.009</td>
<td>-.031 to .050</td>
<td>.000</td>
<td>22.42</td>
<td>15</td>
<td>-.051 to .117</td>
<td>17.585</td>
<td>.001</td>
<td>.669</td>
</tr>
<tr>
<td>Institutional groups</td>
<td>.061</td>
<td>.061 to .061</td>
<td>.000</td>
<td>3.11</td>
<td>5</td>
<td>-.044 to .078</td>
<td>5.382</td>
<td>.001</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note.—Correlations and variances rounded to three decimal places.

*a Frequency-weighted mean correlation.

*b Variance explained by sampling error.

*** $p < .001$. 
to a corrected variance of .015 for the data as a whole. Although variance explained by sampling error for the subgroups did not reach 75 percent (suggesting the presence of additional moderators), type of sample did appear to moderate the follow-ups/repeated contacts-response rate relationship. For stamped/metered return postage, the subgroup correlations were .009 (consumer groups) and .061 (institutional groups), the latter one was significantly different from zero, and the average corrected variance in the subgroups was .000 as compared to a corrected variance of .001 for the data as a whole. In the institutional (but not in the consumer) subgroup, the variance explained by sampling error exceeded 75 percent. Although weak, type of sample did seem to moderate the stamped/metered return postage-response rate relationship.

Discussion

KEY RESULTS AND IMPLICATIONS

Four key findings can be identified from the meta-analysis results reported in this study. First, the relationships between appeals and response rate, and survey length greater than four pages and response rate, were statistically significant and derived from one population; that is, the relationships are generalizable across the studies investigated. Thus, in designing mail surveys, regardless of the target population, researchers that use a cover letter that includes appeals and a survey of less than four pages should be able to increase their response rates.

Second, the associations between reply/return envelope and response rate, incentives less than or equal to $.50 and response rate, and incentives greater than $1.00 and response rate, were statistically significant but derived from more than one population; that is, the relationships are situation specific and there is a need to examine potential moderators. Because the moderators in this study were not significant in the cases of these three predictors, to increase response rates by using one of these factors, researchers must consider other variables that may influence individuals' willingness to respond (e.g., situational factors, personality characteristics).

Third, for the remaining predictor-response rate relationships investigated, the correlations were not statistically significant, and subgroup correlations identified by moderators need to be examined. Thus, without consideration of other variables that may influence individuals' response tendencies, these factors do not appear to enhance response. In this study, two potential moderators were examined.
The fourth finding pertains to the two moderators, year of publication and type of sample. In terms of the meta-analysis results for the year of publication moderator, several key findings can be specified. First, the relationship between preliminary notification and response rate varied across time periods, and for more recent times (1976–87), the correlation was statistically significant. Preliminary notification has been most effective in recent time periods, suggesting that respondents' changing attitudes are involved. Current researchers should therefore continue to use preliminary notification to enhance response rates. Although time or cohort influences appear to be significant, the results also suggest other moderators, and thus subgroups, could be affecting this relationship. Second, the association between stamped/metered return postage and response rate also varied with time, and all three correlations were statistically significant. Moreover, the results suggest that searching for additional moderators of this relationship may not be necessary. Third, for the remaining predictor-response rate relationships, year of publication was not concluded to be a significant moderator.

Several key findings also can be specified for the meta-analysis results for the type of sample moderator. First, the association between follow-ups/repeated contacts and response rate differed for samples of consumer groups and institutional groups, and the correlation for the latter was statistically significant. Thus follow-ups/repeated contacts seemed to have a greater effect on institutional than consumer groups' response rates, and the designers of mail surveys for these different target groups should be aware of such differential responses. For both subgroups, however, additional moderators may be influencing this association. Second, the relationship between stamped/metered return postage and response rate differed (albeit weakly) for samples of consumer and institutional groups, and again, the latter correlation was statistically significant. There seemed to be a greater effect on the response rates of institutional as compared to consumer groups. However, for the consumer groups, at least, additional moderators may be operating. Third, for the remaining predictor-response rate relationships, type of sample was not concluded to be a significant moderator.

The statistically significant effects in this study were moderate in magnitude. The practical implications of these findings can be illustrated in terms of a binomial effect size display (BESD; Rosenthal 1984; Wolf 1986). The predictor with the largest frequency-weighted mean correlation with response rate was incentives less than or equal to $.50 (r = .184). Using this incentive would increase mail survey response rates approximately 18.4 percent, or from a response rate of about 40.8 percent to 59.2 percent (see Rosenthal 1984, p. 131). Incentives of $1.00 or more would improve response rates 12.2 percent.
This is consistent with the findings of Goyder (1982), Heberlein and Baumgartner (1978), and Yu and Cooper (1983).

Preliminary notification for studies conducted between 1976 and 1987 and follow-ups for institutional respondent groups were also significantly related to response rates. All previous meta-analyses, with the exception of Armstrong and Lusk (1987), who did not examine any type of contact, report similar findings. The present results suggest that preliminary notification (1976–87) increased response rate by 28.5 percent, while follow-ups for institutional groups increased response rate by 30.6 percent.

Like Fox, Crask, and Kim (1988) and Armstrong and Lusk (1987), in the current study, postage was an important predictor of response rate. For all time periods, stamped/metered return postage was significantly related to response rate. Most recently (1976–87), stamped-metered postage increased response rate 2.4 percent, while in the past (1940–70), the enhanced effect on response rates was even greater (6.6 percent). The negative effect obtained for the period 1971–75, however, may be a statistical anomaly due to the small number of observed correlations (5). On the other hand, the trend over the 47-year time period indicates that the affect of postage is declining and may signal an expectation on the part of respondents that postage is a survey requirement and its absence may only tend to deter rather than motivate response. Moreover, for institutional samples, stamped/metered return postage increased response rates about 6.1 percent. Again, however, this result is based on a small number of correlations (5).

Compatible with the work of Heberlein and Baumgartner (1978), results from the current study indicated that response rates would be reduced by 7.8 percent for survey lengths of greater than four pages. In contrast to Yu and Cooper (1983), appeals (4.7 percent) and reply/return envelope (7.9 percent) had a positive effect on response rates in the present investigation. Overall, the average effect size across all predictors of response rate ($r = .065$) in the current study would indicate an increase in response rates of about 6.5 percent, or from 46.75 percent to 53.25 percent (see Rosenthal 1984, p. 131).

Two current findings were not consistent with previous research. While Fox, Crask, and Kim (1988), Goyder (1982), and Heberlein and Baumgartner (1978) reported a significant effect for sponsorship, the present results did not support this finding. One reason for this difference may stem from the level of aggregation of studies used in the present study. By aggregating groups within the institutional category, subpopulation effects may have been obscured. For instance, Heberlein and Baumgartner (1978) found that government sponsorship increased response, while this study was unable to investigate this issue because only six data points were available for analysis. Also,
Yu and Cooper (1983) reported a significant relationship between non-monetary incentives and response rate. Nonmonetary incentives were not a significant predictor of response rates in the present study, although the overall correlation did indicate the presence of a positive effect on response. This result as well as those discussed earlier may diverge from the present results, in part due to the inclusion of both more recent studies and a greater number of effect sizes than in past work.

LIMITATIONS AND FUTURE WORK

Nevertheless, the results of this investigation are generally compatible with and extend those of prior review (e.g., Harvey 1987; Kanuk and Berenson 1975; Linsky 1975) and meta-analysis (e.g., Fox, Crask, and Kim 1988; Goyder 1982; Heberlein and Baumgartner 1978; Yu and Cooper 1983) articles about mail surveys. The extensions involve the clarification of prior work by examining a greater number of factors that influence response rate through a large number of study outcomes and the investigation of moderators. The current study, however, is not without limitations, and thus, several directions for future research can be suggested.

First, the moderators that were chosen for investigation did not clarify the relationships as much as had been expected. Perhaps this was due to a poor coding scheme or perhaps the choice of moderators which were too global in nature. For instance, to ensure an adequate number of correlations for meta-analyses of individual predictors, it was necessary to combine educational, industrial, health care, governmental, and other institutional samples into one subgroup. This combination may have obscured or perhaps cancelled important differences in results among the groups. With a few exceptions, predictor-response rate relationships appear to be moderated by other variables. Future work concerning moderators might focus on characteristics of respondents (e.g., personality, prior knowledge, involvement, expertise), additional details about samples (e.g., academic, government, general public), and the situations (e.g., work vs. home or on-site intercepts) in which surveys are administered and/or completed. Also, results from census pretesting (Riche 1987) indicate there are significant geographic differences in response rates due to regional as well as metropolitan versus rural respondents. These results as well as those from BLS surveys (Riche 1987) and the Goyder (1982) study indicate that other potential moderators to be examined may relate to cultural and subcultural (e.g., blacks, Hispanics) differences in response to surveys.

Second, the predictors of response rate were investigated in a limited way in this study; in general, a simple "yes/no" coding scheme was
used. Along the lines of the work of Worthen and Valcarce (1985) and Armstrong and Lusk (1987), future studies could investigate each predictor in this study in greater detail as well as focus on other and multiple predictors of response rate simultaneously. Likewise, some of the findings from the current research may be unstable due to the small number of effect sizes that were available to be synthesized (e.g., appeals, deadlines, incentives greater than $.50 and less than or equal to $1.00). As such, additional primary research on these predictors of response rate is necessary. In these ways, more comprehensive meta-analyses could be conducted in future work.

Third, because the 184 data points (study outcomes) for the meta-analysis were obtained from 115 studies (articles), an average of 1.6 per study, there is nonindependence of study outcomes in several cases. For any given predictor of response rate, however, only one effect size per study was used in the meta-analyses. Although nonindependence may not be of concern when an individual factor influencing response rate is examined, it affects the overall meta-analysis encompassing all predictors. To address this issue, the average correlation per study across factors manipulated in that study was used in the overall meta-analysis. In future work, various statistical adjustments to the correlations or limiting the number of correlations included from any one study could be investigated as alternatives to the approach used here. Likewise, whether more than one manipulation (e.g., personalization and return postage) for the same subjects in one study reduces or enhances response rates could be investigated in future work.

Fourth, a "file drawer problem" may exist in that inclusion of missing studies could alter the results of this investigation (Rosenthal 1979). The inclusion of unpublished studies, especially in cases where results are based on only a few published studies or the confidence intervals are close to zero, could affect the findings reported here. In future investigations, researchers might include both published and unpublished work in meta-analyses to obtain more comprehensive findings.

Fifth, although guided by prior theoretical work, this investigation did not present a complete conceptualization or test of how and why people respond to mail surveys. Given the results of this investigation, future work could focus on testing, with meta-analysis as well as new data and experimental studies, a comprehensive model of mail survey response behavior. Candidates for the comprehensive theory include such social psychological theories as attribution and self-perception theory (cf. Hansen 1980; Reingen and Kernan 1977, 1979) and dissonance theory (Furse and Stewart 1982; Hackler and Bourgette 1973) as well as the equity/exchange model (Childers 1986; Childers and Skinner 1985). In the latter, the three constructs of cooperation, com-
mitment, and trust are integrated to provide an explanation for the impact of such significant effects as preliminary notification, follow-ups, and appeals (increasing respondent outcomes) and postage/return envelope and questionnaire length (reducing respondent inputs) as well as incentives and sponsorship (increasing respondent trust).

Despite some limitations, the current research incorporated several aspects of a properly conducted meta-analysis as identified by Bangert-Drowns (1986) and Bullock and Svyantek (1985). Moreover, meta-analysis procedures of Glass (1977), Hunter and Schmidt (1990), and Hunter, Schmidt, and Jackson (1982) were combined in this study. The results of this research, in conjunction with those from the suggested future research efforts, should clarify understanding of mail survey response behavior and, thus, permit the design of surveys which increase response rates. The current investigation was intended as a step in that direction.

Appendix
Studies Included in Meta-analysis


A Meta-analysis of Mail Surveys


References


