The Social Desirability Variable in Organizational Research: An Alternative Explanation for Reported Findings

KENNETH W. THOMAS
University of California, Los Angeles

RALPH H. KILMANN
University of Pittsburgh

Because questionnaires are quite prevalent in management research, awareness of the effects of social desirability is critical. Data on measures of conflict behavior are presented to illustrate two potential spurious effects—elevation of means and misleading correlations. General substantive and methodological implications for management research are discussed.

Empirical studies reported in management and organizational theory journals have relied heavily upon questionnaires—managers' ratings of organizational phenomena and/or pencil-and-paper measures of personality. In the few cases where objective measures have been utilized (e.g., profits, promotions, production), they usually have been studied in relation to questionnaire data. A quick survey of the offerings in the Academy of Management Journal from September 1973 through June 1974, for example, indicated that 89 percent of the empirical studies used questionnaire data. Fifty-five empirical studies were reported, compared with 17
literature reviews, commentaries, and theoretical pieces. Of these 55, 49 reported using questionnaires to gather data.

A vast number of studies rely on the use of these subjective instruments; thus one might expect that the reliability and validity of such instruments have received considerable investigation, or else that they are used with great caution and tentativeness. Neither of these seems to be the case, however, in most of the research studies reported. Instead, instruments designed by the researcher for a particular study often are utilized without extensive pretesting, and researchers often use existing instruments because "others have used them."

There are reasons for this. An investigator primarily interested in empirical, substantive research may not expect or want to spend time developing instruments before he can get on with his work. Consequently, he often is willing, as suggested, to use whatever instruments are available and to let the psychometricians and the instrument developers worry about the quality of instruments in a particular area of study. Moreover, observing some of the guidelines for instrument development (1, 15), it is not surprising that many individuals do not engage in this process, for it generally requires a considerable time investment (several years, perhaps) to develop a high quality measure. When the instrument has reached this stage, moreover, there is no guarantee that anyone but the developer will actually use the instrument; therefore, it becomes a fairly high risk strategy to engage in rigorous instrument development.

The dangers involved in using untested instruments have been voiced on many occasions elsewhere in the literature. However, it is not the purpose here to generate more methodological guilt and anxiety either for the present authors or for their colleagues. Rather, the intent is to draw attention to a specific source of methodological concern which is felt to have significant implications for management research and which, until Golembiewski and Munzenrider's recent work (13), has not received attention in the management literature. This source of concern is the social desirability variable (9). The intent here is to make organizational researchers more sensitive to the social desirability of the questionnaire items they use and to its implications for results, so that they can avoid some important pitfalls in the collection and interpretation of questionnaire data. It is also hoped that more substantive and methodological research may be stimulated on this topic.

This article attempts to illustrate the importance of the social desirability variable in organizational research by focusing on conflict research as a case in point. The potential biasing effects of social desirability on measures of conflict-handling behavior will be empirically examined, and the results will be used to examine the validity of some previous conclusions from research on organizational conflict.
SOCIAL DESIRABILITY ISSUES IN THE ASSESSMENT OF CONFLICT BEHAVIOR

In the past few years, a five category scheme for interpersonal conflict-handling behavior has gained some prominence in organizational research. First introduced by Blake and Mouton (3), this scheme appears to represent a significant improvement over the simpler cooperative-competitive dichotomy. As interpreted by Thomas (26), this newer scheme combines two separate dimensions: cooperativeness (attempting to satisfy the other party's concerns) and assertiveness (attempting to satisfy one's own concerns). These two dimensions combine to define five conflict-handling modes: competing (assertive, uncooperative), collaborating (assertive, cooperative), avoiding (unassertive, uncooperative), accommodating (unassertive, cooperative), and compromising or sharing (intermediate in both cooperativeness and assertiveness).

Empirical research using these five conflict-handling modes has tended to rely on questionnaire data. These studies have indicated that collaborating is reported as the most prevalent mode (4, 18, 19, 27), and that collaborating is linked to a number of "positive" outcomes—successful organizational performance (18, 19), positive regard by peers (25), productivity of decision making between supervisors and subordinates (4, 23), individual self-actualization (2), and a variety of other functional states (24).

Thomas (25) suggested that some of these findings might be an artifact of the social desirabilities of the five conflict-handling modes. First, the relative magnitude of the average scores on the five conflict-handling modes may be simply a reflection of the social desirability of the questionnaire items. In research with the social desirability variable, Edwards (12) has consistently found correlations above .80 between the social desirability of questionnaire items and the proportion of subjects who endorse the items as describing themselves. These correlations have not been reduced by anonymity (9), and they have been high in ratings of others (10). A priori, notions of social desirability would be expected to be strong in the area of conflict where society has a vested interest in developing norms of acceptable and unacceptable behavior. Second, the possibility that some conflict-handling modes are regarded more positively than others raises the issue of the "halo effect." Some of the findings cited above may represent simply a tendency to associate "good" conflict behaviors with other "good" variables in an individual's ratings.

The present study attempted to shed some light on these possibilities by examining relevant properties of three conflict-handling mode instruments—those designed by Blake and Mouton (3), Lawrence and Lorsch (18), and Hall (14). Specifically, the study examined: (a) the social desirabilities of the items describing the five conflict-handling modes in these instruments, (b) the relationship between these social desirabilities and subjects' self-report scores on the conflict-handling modes, and (c) the relationship be-
between self-assessment of socially desirable conflict modes and self-assessment on other desirable qualities.

**PROCEDURE**

The study used 115 students in three sections of a graduate course in Behavioral Science for Management at the University of Pittsburgh. Eighty-six subjects completed packages of instruments consisting of the conflict-handling mode instruments arranged in random order, followed by a set of response style measures in random order which included the Crowne-Marlowe Social Desirability Scale (6) and the Edwards Social Desirability Scale (11). The remaining 29 subjects rated the individual items in each of the conflict-handling mode instruments from “1” to “9” according to their social desirability, using the procedures adopted by Edwards (8).

In order to make the three conflict instruments more comparable, slight changes were made in the instructions for two of the three instruments. Subjects were asked to **rank** the five Blake-Mouton conflict statements from most to least typical as descriptions of their own behavior, and they were instructed to rate each of the 25 Lawrence-Lorsch proverbs on the extent to which it described their own approach to disagreements, using Burke's (4) modification of the Lawrence-Lorsch response categories. No changes were necessary for the Hall instrument.

**RESULTS AND DISCUSSION**

**Social Desirability of the Modes**

Table 1 shows means and standard deviations of subjects' average ratings of social desirability for the items representing each conflict-handling mode in the three instruments. Some modes clearly are seen as more desirable than others within each instrument. A one-way analysis of variance with repeated measures on subjects (28) yielded F values (with 4, 112 d.f.) of 51.2 for the Blake-Mouton modes, 106.2 for the Lawrence-Lorsch modes, and 32.0 for the Hall modes. All are significant at \( p < .001 \).

Moreover, there is some consistency in the ordering of modes by social desirability across the three instruments. The ranks of the mean social desirability ratings for the five conflict modes within each instrument are shown in parentheses in Table 1. The three sets of ranks yield a Kendall (17) coefficient of concordance of .82 \((p < .05)\). Collaborating is rated as the most desirable mode on all three instruments. This finding is consistent with Lawrence and Lorsch's statement that managers reported collaborating (confrontation) "was the ideal way in which conflict should be resolved" (19, p. 74). In contrast, compromising is average to above average, accommodating mixed, competing average to below average, and avoiding below average.
TABLE 1

Mean and Standard Deviations of Subjects' Average Social Desirability Ratings of Mode Items for the Three Instruments (N = 29)

<table>
<thead>
<tr>
<th>Conflict-Handling Modes</th>
<th>Instruments&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blake-Mouton</td>
</tr>
<tr>
<td>Competing (Forcing)</td>
<td>3.90 (4)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Collaborating (Confrontation)</td>
<td>7.90 (1)</td>
</tr>
<tr>
<td>Compromising (Sharing)</td>
<td>7.38 (2)</td>
</tr>
<tr>
<td>Avoiding (Withdrawal)</td>
<td>3.76 (5)</td>
</tr>
<tr>
<td>Accommodating (Smoothing)</td>
<td>5.52 (3)</td>
</tr>
<tr>
<td></td>
<td>1.88</td>
</tr>
</tbody>
</table>

<sup>a</sup> For this analysis, individual data for the Lawrence-Lorsch and Hall instruments consisted of a subject's average rating of social desirability over the 5 or 12 items describing each mode. The Blake-Mouton instrument has only one item per mode.

<sup>b</sup> Numbers in parentheses are ranks of conflict mode means within instruments.

Relationship Between Social Desirability and Self-Assessment

Mean self-assessment scores were calculated for the five conflict-handling modes on each of the three conflict instruments. These scores were found to vary closely with the social desirabilities of the five modes on all three instruments. The Pearson correlations between social desirability means and self-report means over the five conflict modes were .94 for the Blake-Mouton instrument (<i>p < .05</i>), .96 for the Lawrence-Lorsch (<i>p < .01</i>), and .98 for the Hall (<i>p < .01</i>). These are Pearson correlations, N = 5 modes, with one-tailed significance tests. The sign on the Blake-Mouton correlation was changed to indicate the positive relationship between social desirability and high assessment (low ranks). The correlations indicate that on the average one can account for over 90 percent of the variation in this sample's aggregate ratings over conflict-handling modes solely in terms of the social desirability of the questionnaire items used to assess them.

To some extent, these correlations may reflect a substantive relationship between the social desirability of the conflict-handling modes and their actual occurrence. That is, there may be some tendency for people to behave in ways which society encourages in conflict situations. However, one would not expect that dynamic to be strong enough to account fully for these correlations. Conflict-handling behavior is influenced by a number of factors besides values and social pressure: by personal needs, conflict of interest, stakes, relative power, the behavior of the other, etc. (26). The construct of a social desirability response set remains a compelling alternative explanation of these correlations—that is, that self-ratings are shaped by the social desirabilities of the items on the instruments.
This explanation also is consistent with other relationships in the data. In both the Lawrence-Lorsch and Hall instruments, different phrasings of the same conflict-handling mode elicit different responses depending on the social desirability of those phrasings. (The Blake-Mouton instrument is not relevant to this analysis, since it contains only one item per conflict-handling mode). Pearson correlations were calculated between mean desirability ratings and mean self-assessment ratings for the set of statements representing each of the five conflict-handling modes on these two instruments. As shown in Table 2, the correlations within all 10 sets of statements are in the predicted direction (positive) and nine of the 10 achieve significance at the .05 level or better.

**TABLE 2**

Correlations Between Mean Social Desirability Ratings and Mean Self-Assessment Ratings for the Set of Items Representing a Given Mode

<table>
<thead>
<tr>
<th>Modes</th>
<th>Lawrence-Lorsch (N = 5 items/mode)</th>
<th>Hall (N = 12 items/mode)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competing</td>
<td>.93**</td>
<td>.91***</td>
</tr>
<tr>
<td>Collaborating</td>
<td>.92*</td>
<td>.51*</td>
</tr>
<tr>
<td>Compromising</td>
<td>.71</td>
<td>.80***</td>
</tr>
<tr>
<td>Avoiding</td>
<td>.90*</td>
<td>.69**</td>
</tr>
<tr>
<td>Accommodating</td>
<td>.91*</td>
<td>.52*</td>
</tr>
</tbody>
</table>

*p < .05, one tail

**p < .01, one tail

***p < .001, one tail

Although the present data are restricted to self-assessment, Edwards' (10) work indicates that the strong correlation between item desirability and ratings also is present in subjects' ratings of acquaintances. In studies using the Interpersonal Check List, Edwards found that the correlation between the social desirability values of items and the frequency of endorsement of items was as great on the whole in subjects' ratings of acquaintances as in ratings of self. The correlations were .82 for males and .84 for females in self-report, and .84 for males and .86 for females in ratings of acquaintances.

Extrapolating from the present findings and those of Edwards, it appears that the social desirability response set provides a reasonable alternative explanation for the relative strengths of ratings of the five conflict-handling modes. The implication is that the relative endorsement of different conflict items by a group does not necessarily reflect the relative frequency of the actual behaviors those items were designed to assess. When subjects using the Lawrence-Lorsch items to rate conflict behavior in their organizations show a higher mean rating for collaborating (confrontation) than other modes, it does not appear justifiable to conclude that those organizations
in fact "used confrontation more than other modes . . . " (18, p. 43). Also, subjects' ratings of confrontation as most typical in disagreements with supervisors do not justify the conclusion that "the most common method is confrontation . . . " (4, p. 397). Subjects' relative ratings of conflict behaviors may be influenced by both the relative social desirability of the items and the relative occurrence of the actual behaviors. To be able to draw conclusions about the relative occurrence of actual behaviors, one would need either a more objective measure of behavior or a rating procedure which controls for social desirability and other possible response biases.

Correlations with Ratings of Other Desirable Characteristics

The 86 subjects who rated themselves on the three conflict behavior instruments also completed the Edwards Social Desirability Scale (11) and the Crowne-Marlowe Social Desirability Scale (6). Both instruments measure the frequency with which subjects endorse a variety of socially desirable statements on different topics as descriptive of themselves.

Indices were calculated of the social desirability of an individual's self-ratings on each of the three conflict instruments. An individual's responses were first standardized about his own mean to control for individual tendencies to use different portions of the response scales. These standardized scores on each item were then multiplied by the social desirabilities of the conflict items to provide an index of the extent to which the subject gave relatively high endorsements to the more desirable conflict items. Pearson correlations between the Edwards and Crowne-Marlowe social desirability scales and these social desirability indices for the three conflict instruments are shown in Table 3. All six correlations are in the expected direction, five of them statistically significant. Thus, the social desirability of subjects' self-ratings on conflict-handling behavior had some tendency to vary with self-ratings on other desirable characteristics.

TABLE 3

Pearson Correlations of Two Social Desirability Scales with Indices of the Social Desirability of Subjects' Self-Ratings on the Three Conflict Instruments (N = 86)

<table>
<thead>
<tr>
<th>Conflict Instruments</th>
<th>Edwards</th>
<th>Crowne-Marlowe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blake-Mouton*</td>
<td>.28**</td>
<td>.26**</td>
</tr>
<tr>
<td>Lawrence-Lorsch</td>
<td>.27**</td>
<td>.23*</td>
</tr>
<tr>
<td>Hall</td>
<td>.42***</td>
<td>.14</td>
</tr>
</tbody>
</table>

* Since low ranks indicate high frequency on the Blake-Mouton items, a low index score indicates relatively socially desirable ratings. For the sake of comparability, the signs of the Blake-Mouton correlations therefore have been reversed in this table.

*p < .05, one-tail

**p < .01, one-tail

***p < .001, one-tail
These correlations clearly are susceptible to explanation in terms of a halo effect in self-ratings, a tendency for different desirable characteristics to vary together in a subject's ratings of a target—in this case, himself. In the case of ratings of others, Edwards (10) has found that an individual's ratings of another person on various socially desirable traits vary together, depending upon the rater's attraction to the ratee. When subjects liked the target person, endorsements of items correlated very highly with the social desirabilities of those items: .93 for men raters and .95 for women. When raters disliked the target person, these correlations decreased to —.38 and —.13.

Again extrapolating from these results and those of Edwards, it appears that the halo effect provides a reasonable alternative explanation for a significant degree of correlation between ratings of conflict-handling modes and other desirable (or undesirable) variables. This possibility implies that the conclusions of several previous studies be regarded with some suspicion.

For example, in his use of the Lawrence-Lorsch items to study conflict-handling modes between managers and their supervisors, Burke (4) asked subjects to rate the modes which occurred in their relationship with their supervisor, as well as the constructiveness with which conflict was handled. Constructiveness of conflict-handling, a desirable state of affairs, was found to correlate positively with what the present study shows to be the two most socially desirable modes on the Lawrence-Lorsch instrument—collaborating and accommodating—and negatively with the two least desirable modes—competing and avoiding. The moderate size of Burke's correlations (the highest was .26) is comparable to the correlations in the present study between the Lawrence-Lorsch scores and the two social desirability scales. In the present study, for example, collaborating correlated .35 with the Edwards Social Desirability Scale and .24 with the Crowne-Marlowe. Since the halo effect constitutes an alternative explanation for these correlations, it does not appear justifiable to accept them at face value as reflecting objective relationships in the phenomena of conflict management.

In a more recent use of the Lawrence-Lorsch items, Ryan and Clemence (24) asked organizational employees to rate their organizations on conflict-handling behavior as well as 24 "organizational effectiveness" variables. A factor analysis of the Lawrence-Lorsch items was performed. The "problem solving" factor correlated positively with each of the desirably phrased (organizational effectiveness) variables and negatively with each of the undesirably phrased (organizational ineffectiveness) variables, attaining significance for 21 of the 24 relationships. Also, the Ryan and Clemence "adversary resolution" factor correlated with each of these variables in the opposite direction, attaining significance for 18 of the 24 relationships. Correlations again were moderate in size. An efficient alternative interpretation of these results is that ratings of effectiveness and conflict variables were mutually responsive to the employee's general evaluation of the organization.
Using the Ryan-Clemence factor loadings to weight the social desirability values of the Lawrence-Lorsch items which were obtained in the present study, the social desirability value of their problem solving factor is relatively high (7.17); the adversary resolution factor is relatively low (4.25). It does not appear justifiable to accept the correlations as representing a substantive relationship between conflict behavior and organizational effectiveness.

One way of minimizing spurious correlations between the conflict instruments and other variables, of course, is to obtain independent and/or objective measures of those other variables. Even here, however, there may be some problem involving social desirability. Ratings of self or others still are likely to vary in social desirability with the rater's general like/dislike of self or others. Variables which affect this degree of like/dislike therefore may be spuriously correlated to socially desirable ratings of conflict-handling behavior. For example, Lawrence and Lorsch (19) found that collaboration was rated as more frequent in high performing organizations than in low performers where performance was determined by a variety of financial data. One factor in this relationship may have been that executives in the top performing organizations felt more positively toward their organization and each other and therefore rated conflict behavior in their organization in more positive or desirable terms.

**GENERAL IMPLICATIONS**

A number of findings and interpretations from previous studies of conflict-handling modes have been noted which are suspect because their results can be explained by the social desirability dynamics noted above—the tendency of means to reflect the social desirability of items, and the tendency for ratings of socially desirable items to vary together. Although data in the present study have focused on self-assessment and on conflict-handling behavior, the authors have cited the research of Edwards to support the generalizability of their results to individuals' ratings of others and to research in other substantive areas. These general dynamics and distortions would be expected to operate in studies of organizational climate, leadership, risk-taking, etc.—anywhere, in short, where ratings are used to assess variables with evaluative overtones. For example, Golembiewski and Munzenrider (13) recently have found a significant relationship between managers' scores on the Crowne-Marlowe Social Desirability Scale and their tendency to rate their own organization as "System IV" on Likert's (20) Profile of Organizational Characteristics.

Although it was suggested earlier that other types of instrument problems (i.e., other aspects of reliability and validity) could undermine the substantive findings of empirical research, this paper has focused on social desirability and, consequently, the discussion of general implications will be limited to this issue.

The first set of implications is concerned with social desirability as a methodological problem. The findings of this study suggest that researchers
should pay more attention to the social desirability variable, its possible biasing effects, and its candidacy as an alternative explanation for their findings. In particular, they suggest that researchers discount high ratings for socially desirable items and be suspicious of correlations between socially desirable variables. In terms of research design, the results also underscore the desirability of obtaining observational or other objective measures where possible to reduce the intrusion of social desirability into the data, or at least of obtaining ratings of different variables from independent sources to minimize halo effects.

Questionnaires remain an efficient and economical way of gathering data, however, and the results also suggest the need to improve the ability to understand and control the effects of social desirability in questionnaire ratings. As a first step, this would require the development of a "nomological network" (5) of the factors which influence the social desirability of ratings. At present, understanding and technology are rather primitive in this area. For example, it is now common practice to use only the Crowne-Marlowe social desirability scale to control for social desirability response sets in the design of questionnaires. However, the social desirability of ratings is not determined by a single dynamic. The two social desirability scales used in this study—measures of the extent to which an individual tends to describe himself in socially desirable terms—show a relatively small intercorrelation. The correlation in the present data between the Edwards and Crowne-Marlowe measures was .21; Crowne and Marlowe (6) report a correlation of .35. Construct validation of the Crowne-Marlowe scale (7) indicated that it assesses a need for approval, but an inspection of items suggests that the Edwards scale may tap self-esteem in addition. Clearly, these would represent two very different sources of social desirability biases in self-report, and there may well be others. Kasl and French (16) concluded that positive self-ratings could represent a defensive maneuver by individuals with a strong fear of failure. Clearly, more research is needed to explore these various sources and their implications for questionnaire design.

A second set of implications concerns the potential substantive implications of social desirability for the practice of management. Here an extrapolation of sorts is made from social desirability to the general issue of evaluative tendencies. Consider that few descriptive phrases in the vocabulary tend to be neutral: in analyzing personality instruments, for example, Edwards (12) found that the items had a bimodal distribution in terms of social desirability—i.e., they tended to be either positive or negative. Consider further that semantic differential research has shown that more variance in ratings is accounted for in terms of an evaluative dimension (good-bad, desirable-undesirable) than any other (22). And, finally, consider the pervasiveness of halo effects in managers' ratings of each other and their organization. The general desirability/evaluation dimension of managers' views of themselves, others in their organization, and the organization itself appears to be a central part—perhaps the most important part—of their
phenomenal field in the organization. In addition to devoting energy to minimizing the intrusion of this influence into ratings of more "objective" phenomena, it also may be productive to examine in detail the sources and consequences of these perceptions of desirability in organizations. For example, consider an organizational climate in which individuals view themselves, each other, and the organization in generally undesirable terms. Such a climate probably would be characterized by cynicism, complaining, and defensiveness. Contrast this with a climate in which individuals perceive selves, others, and the organization in generally desirable terms. Here, one might expect to find pride, mutual admiration, and enthusiasm.

It is suggested that a major component of organizational leadership involves enabling the organization's members to perceive themselves, each other, and the organization in positive ways. Leaders following this approach might devote effort to directly influencing workers' perceptions of desirability—for example, by demonstrating the worth of the organization's mission and activities to its members, emphasizing the importance and value of personal contributions, building pride in a work unit's abilities, and placing negative events in perspective. Investigating the activities which leaders use to make members feel positive about themselves, each other, and the organization may well be another step towards understanding effective leadership. Although this strategy sounds similar to what practitioners refer to as "building morale" or as the harnessing of Maslow's "ego needs" (21), this seems to be an important component of leadership activity which has received little attention in the recent management literature.

REFERENCES

25. Thomas, K. W. *Conflict-Handling Modes in Interdepartmental Relations* (Doctoral dissertation, Purdue University, 1971).