Self-Reports in Organizational Research: Problems and Prospects

Philip M. Podsakoff
Dennis W. Organ
Indiana University

Self-reports figure prominently in organizational and management research, but there are several problems associated with their use. This article identifies six categories of self-reports and discusses such problems as common method variance, the consistency motif, and social desirability. Statistical and post hoc remedies and some procedural methods for dealing with artifactual bias are presented and evaluated. Recommendations for future research are also offered.

A casual inspection of published research in organizational behavior or management shows the self-report to be well-nigh ubiquitous as a form of data collection (cf. Dipboye & Flanagan, 1979; Gupta & Beehr, 1982; Mitchell, 1985; Sims, 1979). Coincident with ubiquity, however, is the apparently widespread suspicion that self-report methodology is the soft underbelly of the organizational research literature (cf. Campbell, 1977). It seems that organizational researchers do not like self-reports, but neither can they do without them. On the assumption that self-reports are here to stay, a dispassionate assessment of this research procedure is needed.

Concern about the problems of self-report research is—or at least should be—shared by practitioners in management or staff positions. Many of the findings disseminated to the management community draw from self-report research. In addition, managers rely heavily on their own self-report research in employee opinion surveys, program evaluation, and human resource planning.

The authors have themselves struggled over the years with the dilemma posed by use of self-reports in organizational research. From the beginning, we were aware that the questionnaire, at best, provides "soft" data, perhaps better than mere opinions with no data at all, but vastly inferior to most other kinds of data. The problem is more serious in some instances than in others, at times so serious that findings are rejected out of hand. But many organizational research issues stubbornly resist reformulation in terms of other approaches. Those who recommend carrying the research into the laboratory and manipulating the variables of interest do, of course, have the textbooks of scientific method on their side, but

---
This article was moved ahead in the publication sequence in order to pair with the earlier accepted Kemery and Dunlap piece which immediately precedes it.

 Portions of this article are based on presentations by William E. Scott, Jr., and Philip M. Podsakoff at the 1985 meetings of the Midwest Academy of Management and by Dennis W. Organ at the 1985 national meetings of the Academy of Management.

 Address all correspondence to Philip M. Podsakoff, Department of Management, School of Business, Indiana University, Bloomington, IN 47405.

Copyright 1986 by the Southern Management Association 0149-2063/86/$2.00.

531
such a strategy often seems to change profoundly the very nature of the issue that prompted our interest in the first place.

Therefore, like many researchers, we have found ourselves in the position of having to use self-reports while knowing that this invited serious problems. The sensible approach seemed to be to either reduce the magnitude of the problem as much as possible, or, failing that, to assess the problems inherent in the data. But no structured guide exists for using self-reports in the most defensible fashion. In our own research, we have tried a number of devices in more or less trial-and-error style, seeking as we went along a more clearly defined notion of just what the relevant problems were, how to relate them to the broader context of research problems, how to measure them, and how to deal with them. We believe that an account of our odyssey, however insufficient as a researcher’s guide, may nonetheless begin constructive discussions of the issue.

First, we provide a rough ordering of the occasions or purposes for which self-reports are used in organizational research, and note the circumstances under which such measures present particular problems. Next, we assess methods for mitigating the problems. Finally, some considerations are noted for future use of self-reports in organizational research.

**Uses of the Self-Report**

The uses of self-reports for data gathering break down roughly into the following categories:

1. Obtaining demographic or otherwise factual data (such as age or sex of respondent, years of tenure, etc.), that are, in principle, verifiable from other sources.


3. Gathering personality data (trait anxiety, need for achievement, locus of control, and so forth).

4. Obtaining descriptions of a respondent’s past or characteristic behavior (e.g., asking supervisors about their “structuring” behaviors), and/or seeking respondents’ intentions of future behavior (e.g., to quit), or how they would behave under certain hypothetical conditions (i.e., various role-playing exercises).

5. Scaling the psychological states of respondents, such as job attitudes, tension, or motivation.

6. Soliciting respondents’ perceptions of an external environmental variable (the supervisor’s behavior, formalization of organizational processes, climate).

**Problems That Arise**

The categories identified above are not mutually exclusive in every case and probably not exhaustive, but they offer a basis for preliminary discussion. Of the six categories noted, the first two appear to offer the least difficulty. Demographic data obtained by self-reports are usually independently verifiable from other sources; often the data can be checked in the aggregate or against archival information. Moreover, although persons may report erroneous information, the practical problems so posed seldom offset the convenience and economy of col-
lecting such data by self-reports (Gupta & Beehr, 1982). In any case, self-report items designed to gather such data are often included because the commitment has already been made to use self-reports for other types of variables.

Manipulation checks basically provide reassurance to the experimenter and his/her colleagues that the independent variable ‘‘registered’’; otherwise they seldom figure importantly in the conclusions drawn from the research.

Categories 3 through 6 present greater problems. Self-report measures of such variables are, in the strictest sense, not verifiable by other means. There may well be other better measures of such variables, but there is no direct means of cross-validating people’s descriptions of their feelings or intentions. Only in the most limited cases—as when we ask someone how many days he or she has been absent, or how many grievances he or she has filed—can the answers be verified, even in the aggregate.

What soon becomes obvious upon examining Categories 3 through 6 is that when using self-reports to scale such variables, we are generally not asking people to report a specific fact or a finite event. We are asking persons to go well beyond that and to engage in a higher-order cognitive process—a process that involves not only recall but weighting, inference, prediction, interpretation, and evaluation. Many times during a brief interval, we are requiring the respondents to work at a fairly high level of abstraction. Thus, the data we obtain are already quite a few steps removed from the level of discrete stimuli and responses.

Nonetheless, despite the rarified form of response to a scale measuring a personality, environmental, or attitudinal variable, the result is still a measure. If the use of this measure is to correlate it with some other measure derived by some other means, the only real problem is establishing the validity of the self-report measure. In principle, this is true of any measure, and there are a variety of methods by which to adduce evidence of validity. For example, the validity of a self-report measure of job characteristics can be established if it discriminates among officially graded job levels, the training or intelligence of people who perform such jobs, or the ratings of such jobs by independent experts. (On the whole, however, our research community has probably been somewhat negligent in documenting validity evidence of self-report measures, particularly with respect to discriminant validity; Schwab, 1980).

The most severe problems arise when measures of two or more variables in Categories 3, 4, 5, and 6 are collected from the same respondents and the attempt is made to interpret any correlation(s) among them. This is the well-known problem of common method variance (Campbell & Fiske, 1959; Fiske, 1982). Because both measures come from the same source, any defect in that source contaminates both measures, presumably in the same fashion and in the same direction. The problem is not eliminated by evidence of validity. One measure may be shown to correlate highly with alternative indices of its referent and the other measure may likewise be shown to correlate with what it should correlate with. But the correlation between the two measures, when both are obtained from same-source self-reports, need not include any variance common to both of their domains (see Figure 1). Thus, a self-report measure of job characteristics and a self-report measure of job satisfaction may each overlap with the variance in their...
respective domains. But that does not guarantee that the overlap in variance of the measures themselves, when taken from the same sources, includes any of the overlapping variance they share with their referent domains. The correlation could erroneously lead us to infer a substantive relationship.

Thus, the most critical problem in the use of self-reports is identifying the potential causes of artifactual covariance between self-report measures of what are presumed to be two distinctly different variables. When the same persons provide us with self-report measures of two or more different constructs, what could account for any correlations we find, other than a real underlying relationship?

Perhaps the most general problem is the **consistency motif**. That is, given what we know from decades of research on cognitions and attitudes, respondents apparently have an urge to maintain a consistent line in a series of answers, or at least what they regard as a consistent line. This creates problems because, first of all, many people have lay theories of how personality, behavior, psychological states, and organizational environments are interrelated (Eden & Leviatin, 1975; Lord, Binning, Rush, & Thomas, 1978; Phillips & Lord, 1986; Staw, 1975). This would be less of a problem if we were asking people to report discrete events, the perception and recall of which are less vulnerable to distortion. But in self-report research, we often ask for summary judgments. When we do so, we are inviting respondents to array their judgments consistent with prevalent lay theories (e.g., the relationship between leader style and group morale).

Furthermore, research by Chapman and Chapman (1967, 1969), Jenkins and Ward (1965), and Berman and Kenny (1976) shows that even the most astute subjects are prey to illusory correlations when exposed to a series of events. Not only do respondents have lay theories about how organizational phenomena ought to be related, they also overestimate the strength of the empirical relationships they have observed between classes of events. Perhaps this is because they do have lay theories; confirming cases are given disproportionate weight in overall impressions and are the most likely to be recalled.

Although the measures of job satisfaction and job characteristics can both be shown to have validity (in the sense of overlapping with their intended domains), the overlap between the measures themselves does not ensure that the domains of interest are interrelated in a substantive sense.

**Figure 1**
Aggravating the consistency motif problem is the fact that self-report measures of different variables are often found to contain items similar in content. That is, a scale purporting to measure job satisfaction may refer to considerate treatment from the supervisor, and another measure given to the same person may ask questions designed to measure the supportiveness given by the supervisor. Little wonder that respondents will answer consistently to what are essentially the same questions from different scales.

Note that the most frequently studied topics within Categories 3, 4, 5, and 6 contain matters about which organizational participants, as well as researchers, are likely to have fairly strong sentiments. Few people are indifferent to how leaders behave, how the job is designed, or how much stress they experience. How a person reports on such matters can obviously be affected by his or her mood, whatever its cause (e.g., a bad night’s sleep, a cold, rainy weather, tax time approaching). Thus, when we obtain self-report measures from a person at one sitting for several variables, there certainly is the risk that transient mood state will contribute a consistent but artifactual bias across the measures.

Most researchers who contemplate the use of self-reports have heard of the social desirability problem (cf. Arnold & Feldman, 1981; Taylor, 1961; Thomas & Kilmann, 1975), so labeled because questionnaire items may prompt responses that will present the person in a favorable light. To the extent that this problem causes only an upward shift in the distribution of responses, it is not a serious concern, at least in the interpretation of correlations involving the scale. Even if the effect of socially desirable responses were to compress the range of responses around the end of the scale, the damage would occur mainly in the attenuation of correlations (because of restricted variance in scale scores).

Unfortunately, the social desirability problem goes further than merely adding bias to the responses. Not only are some responses to some items more socially desirable than others, certain reasons for responses are also more ego-flattering than others. Thus, suppose I answer a self-report measure of stress by indicating that I experience severe job-related tensions. I am apt to respond to other items that implicate poor supervision, irrational policies and procedures, incompetent subordinates—as opposed, perhaps, to my own inability to work constructively with others or my own lack of planning. Social psychological research in self-attribution (cf. Mitchell, Green, & Wood, 1981) has clearly shown that we are not indifferent to the reasons we ascribe for our failures or our problems.

Answers to two sets of self-report items may also display artifactual covariation because the behavior of responding is, to some extent, under the control of various cues in the stimulus setting. These cues may be quite subtle, even to the extent that the respondent is unaware that they affect his or her behavior. For example, if the subject responds to items in two or more scales while at work in the middle of the afternoon, his or her responses could be influenced by a report he or she has just read or written, a prior conversation, or hunger. The problem here is not the effect of such stimuli on responses to any one scale taken by itself; any measure, self-report or otherwise, will pick up some irrelevant variance. The problem is that such stimuli could affect responses to two or more different scales in the same fashion, resulting in a potentially misleading correlation. And lest the
reader think we have descended to the nittiest nitpicking, remember that our research thrives on correlations in the .20-.40 range, when found statistically significant. Therefore, systematic overlapping variance of 4%—not very much—can give misleading signals.

Remedial Approaches

The array of potential problems we have identified would seem to discourage even the hardiest soul from contemplating a research design involving correlations among self-report measures. But sometimes there is no realistic alternative if we wish to do any research at all on certain questions. This has led to a variety of forms of first aid for patching up some of the damage.

Statistical and Post Hoc Remedies

Some procedures which are used to address the common method variance problem are used after the variables in the study have already been measured. In general, most of these techniques require the use of statistical procedures in an attempt to isolate the covariance due to artifactual reasons.

Harman’s one-factor test. Perhaps the first statistical procedure to be used in an attempt to control for common method variance is Harman’s single-factor test. In this procedure, all of the variables of interest are entered into a factor analysis. Following this, the results of the unrotated factor solution are examined to determine the number of factors that are necessary to account for the variance in the variables. The basic assumption of this technique is that if a substantial amount of common method variance is present, either (a) a single factor will emerge from the factor analysis, or (b) one “general” factor will account for the majority of the covariance in the independent and criterion variables. Representative examples of the use of this procedure can be found in the work of Greene and Organ (1973), C.A. Schriesheim (1979), J.F. Schriesheim (1980), and Podsakoff, Todor, Grover, and Huber (1984).

Although the single-factor test is reasonably straightforward and easy to apply, there are some problems inherent in its use. First, as anyone who has used factor analysis has discovered, the likelihood of finding more than one factor increases as the number of variables increases. Thus, the single-factor test becomes increasingly less conservative as the total number of variables increases.

Second, to our knowledge, no specific rules have been explicated on how many factors the researcher should expect from the factor analytic procedure. Obviously, when only one factor emerges from the analysis, it is quite possible that common method variance accounts for most of the interrelationships. However, it is not clear how many additional factors must be discovered or the amount of variance the first factor must extract before it is identified as a general factor. Another complication is that researchers who have employed this test have often failed to report the amount of variance accounted for by the factors in their analysis.

Few rules have been established to suggest when the procedure is acceptable. For example, J.F. Schriesheim (1980) employed a six-variable subset of the 14 variables used in C.A. Schriesheim’s (1979) earlier study. One might expect that
the most appropriate use of the single-factor test would be of the six variables used in the 1980 study. However, J.F. Schriesheim (1980) reported evidence from C.A. Schriesheim’s (1979) earlier study with the full complement of 14 variables as evidence for a lack of common method variance. Although it is not our intent to criticize J.F. Schriesheim’s (1980) methods, we do note the ambiguity attendant to this procedure.

Remember, too, that factor analysis is “blind” in that it extracts covariance without regard to the reason for the covariance. If four variables are highly intercorrelated because of valid functional relationships among them, one factor might well account for the covariance in the variables. Consequently, there is the chance of throwing out functional interrelationships along with common method variance.

Partial correlation procedure. Another statistical procedure attempting to deal with the common method problem is in some ways an extension of the single-factor test. In this approach, the hypothesis to be tested is whether the relationships among the variables of interest still exist after the common method factor has been statistically controlled. As in the single-factor test, the first step in this technique requires a factor analysis of all the variables in the study. Following this, the first unrotated factor (which is assumed to contain the best approximation of common method variance if it is a general factor on which all variables load) is partialled out and the relationships between the independent and criterion variables are again examined to determine whether any meaningful correlation still exists.

Because it is statistically more conservative, the partial correlation procedure has been considered by some to be a stronger test of the potential biasing effects of common method variance. Despite this potential advantage, however, several problems may also be identified with the use of this procedure. One problem has to do with the lack of knowledge regarding how much of the first factor extracted in the factor analysis contains common method variance. As noted by Podsakoff and Todor (1985), we have no way of identifying which portion of the covariance contained in the first factor is due to functional relationships and which portion is attributable to the use of common methods. Furthermore, as Kemery and Dunlap (1986) show, spurious negative partial correlations may result from this form of analysis. Under these circumstances, interpretation of the findings becomes difficult, if not impossible (e.g., Organ & Greene, 1981).

A final problem with the use of the partial correlation procedure is that it is highly sensitive to the number of variables used in the particular study of interest. Because the structure of the first factor is influenced by the number of variables included in the analysis, researchers with different numbers of variables will partial out differing amounts of variance. Therefore, even if initial correlations among specific variables are the same, their partial correlations may differ considerably.

Elimination of social desirability. Although conceptually distinct from the partial correlational procedure described above, another related procedure involves measuring respondents’ need for social approval on one of the commonly used measures of social desirability (Crowne & Marlowe, 1964; Edwards,
1970), and examining the relationships among the variables of interest with this variable partialed out. Thus, if internal locus of control and job satisfaction are correlated with each other, and each is also correlated with scores on the Crowne-Marlowe scales, then partialling out the latter should remove the common method variance due to social desirability.

Unfortunately, our experience has shown us that most commonly used measures of social desirability are not highly related to other measures in our field. (The correlations are typically in the .15-.30 range.) Therefore, the effects of the partialling procedure are frequently quite nominal (cf. Ganster, Hennessey, & Luthans, 1983). For example, if A and B correlate .30 with each other and each correlates .25 with a social desirability measure, A and B still correlate .25 after partialling out social desirability. Thus, although the social desirability problem threatens much same-source, self-report research, statistical approaches to correcting for this particular problem do not show much promise.

Scale trimming. A final post hoc approach, which is related to the ones described above, may be called scale item trimming. This approach entails the trimming (or elimination) of items that constitute obvious overlap in what are purported to be separate (or distinct) measures. A recent example of this approach has been illustrated by Birnbaum, Farh, and Wong (1986). Aware that previous research using the Job Diagnostic Survey (JDS) to measure job characteristics and the Job Descriptive Index (JDI) to assess employees’ work satisfaction had indicated that these two scales did not possess adequate discriminant validity, Birnbaum et al. used only a subset of the items on the JDI work satisfaction subscale in their study.

The logic of the item trimming approach is quite simple. It assumes that the researcher can identify those items that the respondents perceive as conceptually similar on the scales of interest. We do not believe that such choices would be capricious, but we can envision rather heated debates about the choices which are made. Even if such a selection were made on the basis of more rigorous procedures (e.g., on the basis of those items which load together on the first factor of a factor analysis), one outcome could be that the psychometric properties of the scale are altered—potentially modifying the conceptual meaning of the variables measured. However, this approach is not without some intuitive appeal, and might be a reasonable alternative if a panel of judges agreed on what items were the same and should therefore be eliminated.

**Procedural Methods**

Our analysis has shown that no simple statistical procedure adequately eliminates the problems of same-source variance. Perhaps this is why some behavioral researchers are skeptical of the results obtained from correlational field studies and have suggested that the best solution is to abandon such research and conduct only experimental studies.

Appeal to experimental procedures as a replacement for correlational field studies, however, begs the question of how to deal with same-source variance. Laboratory studies do not lend themselves well to the examination of certain types of behavioral phenomena (interactions among stable group members, long-
term leader-member interactions), and experimental field studies are difficult and expensive to conduct. Moreover, because correlational field studies often provide useful information about relationships among important variables in actual organizational settings, few would advocate that they be totally discarded. Many researchers will still find themselves in situations where it will be absolutely necessary to collect self-report measures from the same subjects, or prohibitively expensive to do otherwise.

**Escalating the unit of analysis.** One solution to this problem may be to escalate the unit of analysis. If a large sample of individuals can be reduced to a smaller but still meaningful number of units (departments, sections) and the variables have meaning at that unit of analysis, the researcher can use a randomly selected portion of respondents within each unit to estimate the value of some of the variables and the remaining persons in the unit for obtaining estimates of the values of the other variables. A recent study reported by Smith, Organ, and Near (1983) provides an example of the use of this technique. These researchers were interested in examining the relationships among leader supportiveness, subordinate satisfaction, and employee citizenship behaviors. Smith et al. (1983) avoided the same-source problem by obtaining measures of satisfaction from half of the respondents in each unit, with peers of these respondents providing the descriptions of the leaders’ behavior. Supervisors rated the citizenship behaviors of the respondents.

Of course, one obvious difficulty with escalating the unit of analysis is that it significantly reduces the sample size to be examined and the power of the statistical test to be used (e.g., from 235 individual respondents to 43 departments). Thus, it requires samples in which meaningful subgroups are large in number. Moreover, in some cases, using a different level of analysis may not be viewed as appropriate. For example, the use of peers of subordinates to assess leader behaviors may appear to fly in the face of some empirical evidence that suggests that leaders respond differently to different subordinates (e.g., Dansereau, Alutto, Markham, & Dumas, 1982; Ilgen, Mitchell, & Fredrickson, 1981; Mitchell, Green, & Wood, 1981; Podsakoff, 1982).

However, in some instances an analysis of the within-groups versus the between-groups variance may be used to determine whether the use of peer reports as measures of leaders’ behavior is justified. If the variance found within the groups is significantly less than the variance found between the groups, one may assume that the behavior of the leaders is perceived as more similar within each group than across the groups. As long as there is no strong theoretical reasoning which would limit the level of analysis at which the phenomenon is to be studied, escalating the unit of analysis does appear to have some advantages for those concerned with the same-source problem.

**Separation of measurement.** In other instances, however, escalating the unit of analysis may not be appropriate to the phenomenon being studied. Under these circumstances, one option may be to collect some of the measures at different times. For example, a researcher interested in the relationship between role ambiguity and job satisfaction may collect the measures of role stress and satisfaction at two different points in time. Better yet, it may be possible to collect some
of the measures in different places (e.g., work vs. home) or by different media (phone survey vs. written questionnaire), or by using some combination of these techniques. This would mitigate the problem of transient mood state and common stimulus cues, and perhaps reduce the effect of respondents’ strain toward consistency.

Scale reordering. A final procedural option would be to alter the design of the questionnaire used to obtain self-reports. Salancik and Pfeffer (1977), for example, have noted that one possible technique that could be used to reduce the effects of consistency artifacts is to reorder the items on the questionnaires such that the dependent or criterion variable follows, rather than precedes, the independent variable. Little research exists to evaluate this strategy, but we feel that the correlations would be similar using either method. However, it may allow us to assess the order effects that exist in the self-report measures.

Where Do We Go From Here?

Despite the problems in the use of self-report measures in organizational research, the practical utility of these types of measures makes them virtually indispensable in many research contexts (cf. Gupta & Beehr, 1982; Sims, 1979). Moreover, there is evidence that under some circumstances self-reports may represent more accurate estimates of population parameters than behavioral measures (Howard, Maxwell, Weiner, Boynton, & Rooney, 1980). Therefore, it is unlikely that such techniques will be abandoned. However, we offer several recommendations regarding the use of such data in organizational research.

First, we strongly recommend the use of procedural or design remedies for dealing with the common method variance problem as opposed to the use of statistical remedies or post-hoc patching up. As is evident from our preceding discussion, one problem common to all of the statistical procedures used to date is that they do not permit one to determine the proportion of covariance between two measures that is attributable to the assessment by the same source. As such, the effectiveness of these techniques for dealing with the problem is, at best, ambiguous. Although each of the procedural remedies requires extra work on the part of the researcher and may necessitate larger sample sizes (e.g., in those studies which escalate the unit of analysis), they do indicate the researcher’s thoughtfulness in attempting to deal with the same-source problem.

We are not saying that each and every researcher should incorporate all of the procedural remedies into his or her design. Nor are we suggesting that all statistical procedures for dealing with the same-source problem be abandoned. If each study used at least one of the procedural techniques outlined in this article, we could determine the amount of agreement across the various methods. Moreover, recent developments in the use of structural equation modeling (LISREL, Joreskog & Sorbom, 1981) may permit researchers to identify better the potential impact of same-source data using statistical procedures. A recent study reported by Glick, Jenkins, and Gupta (1986) illustrates this point nicely.

In their study, Glick et al. (1986) examined the effects of method variance on the relationships between job characteristics and three attitudinal outcomes: effort, general satisfaction, and challenge satisfaction. They accomplished this by
(a) obtaining reports from three separate data sources (interviews of job incumbents, card sorts by job incumbents, and observations by trained observers), and (b) testing the viability of several models of the potential effects of method variance on the job characteristics-outcome variable relationships using structural equation analysis procedures. Glick et al. found that although the relationship between job characteristics and effort and challenge satisfaction existed independent of method effects, almost all of the variance shared between job characteristics and general satisfaction was attributable to common methods—suggesting that the relationships between the job characteristics measures and job satisfaction may indeed be inflated by same-source variance.

It is instructive to note that in order to assess the effects of common method variance in their study, Glick et al. (1986) obtained some of their data regarding the multiple traits from multiple (different) sources. Herein lies an interesting paradox. For, by gathering the independent and criterion variables from different sources, a researcher already has independent assessment of the variables and it would appear no longer necessary to examine the amount of variance attributable to the common methods used. We believe, however, that such studies have great utility for two reasons. First, our understanding of the amount of variance attributable to the use of same-source data is woefully inadequate. Second, recent developments in the use of confirmatory factor analysis techniques (cf. Schmitt & Stults, 1986; Widaman, 1985) to partition the information from multitrait-multimethod matrices into that variance attributable to trait, source, and methods factors appear to offer great promise. Only when we conduct more studies which actually measure the variables of interest with different methods and from different sources are we going to be eventually be able to assess the general effects of same-source variance on the findings obtained.

Finally, our examination of the problems inherent in correlational field studies has made it obvious that we need to develop a more adequate theory of self-report behavior as behavior (Buchwald, 1959). Although the development of such a theory goes beyond the scope of this article, we feel it would necessarily include an analysis of how respondents perceive and record events in the world, as well as how they weight these perceptions, make inferences, interpret questions, and make evaluations. In addition, we need to be able to identify more clearly the factors that control self-reporting behavior, and what is going on when the individual is responding to questions posed by a researcher. Such a theory will prove useful not only in improving methodology, but also to our understanding of the phenomenology of organizational members.

The quality of the knowledge passed on by researchers to our management constituency would be considerably enhanced by progress in resolving the dilemmas in self-report research. Furthermore, such developments could improve the efforts of managers themselves in generating their own site-specific knowledge.

**Summary and Recommendations**

The following recommendations are offered for future research using self-report measures. These recommendations represent an ordering of the procedures that we feel should be used.
1. Ideally, researchers should obtain multiple measures of the conceptually crucial variables from multiple sources using multiple methods. (Conceptually crucial variables are those whose functional relationship is central to theoretical interpretation of the study.) Using structural equation modeling techniques, researchers could assess the relationships among the variables with and without common method variance (CMV), and therefore increase our knowledge of the effects of this phenomenon.

2. If it is not possible to obtain all of the conceptually important variables from multiple sources, researchers should attempt to obtain the independent and criterion variables from different (i.e., procedurally independent) sources, thus avoiding the potential confounding effects of CMV.

3. If conceptually appropriate, data should be aggregated to a larger unit of analysis, using part of each unit to estimate the independent or predictor variable and another part to estimate the criterion or dependent variable. (Even if aggregation is conceptually appropriate, within- vs. between-groups analysis of variance may be necessary to demonstrate that the aggregation process is justified.)

4. If all subjects have to be used for all measures, conceptually crucial variables should be taken at different times, preferably with different scaling formats and in different settings.

5. The scale trimming procedure should be employed to eliminate obvious overlap in items on the independent and criterion measures scales.

6. Failing all else, researchers should at least report results from a test of the single-factor hypothesis as an explanation of the intercorrelation of the variables of interest. The results of such a test cannot be interpreted unequivocally, but may provide a useful note of information to the reader.

References


C.A. Schriesheim (Eds.), Leadership: Beyond establishment views (pp. 81-103). Carbondale, IL: Southern Illinois University Press.


Received January 3, 1985

First revision received November 29, 1985

Accepted April 17, 1986

*Philip M. Podsakoff is an associate professor of personnel and organizational behavior in the Department of Management at Indiana University.*

*Dennis W. Organ is a professor of organizational behavior in the Department of Management at Indiana University.*