Journal of Applied Psychology
2003, Vol. 88, No. 5, 879–903

0021-9010/03/$12.00 DOI: 10.1037/0021-9010.88.5.879

Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies

Philip M. Podsakoff, Scott B. MacKenzie, and Jeong-Yeon Lee
Indiana University

Nathan P. Podsakoff
University of Florida

Interest in the problem of method biases has a long history in the behavioral sciences. Despite this, a comprehensive summary of the potential sources of method biases and how to control for them does not exist. Therefore, the purpose of this article is to examine the extent to which method biases influence behavioral research results, identify potential sources of method biases, discuss the cognitive processes through which method biases influence responses to measures, evaluate the many different procedural and statistical techniques that can be used to control method biases, and provide recommendations for how to select appropriate procedural and statistical remedies for different types of research settings.

Most researchers agree that common method variance (i.e., variance that is attributable to the measurement method rather than to the constructs the measures represent) is a potential problem in behavioral research. In fact, discussions of the potential impact of common method biases date back well over 40 years (cf. Campbell & Fiske, 1959), and interest in this issue appears to have continued relatively unabated to the present day (cf. Bagozzi & Yi, 1990; Bagozzi, Yi, & Phillips, 1991; Campbell & O'Connell, 1982; Conway, 1998; Cote & Buckley, 1987, 1988; Kline, Sulsky, & Rever-Moriyama, 2000; Lindell & Brandt, 2000; Lindell & Whitney, 2001; Millsap, 1990; Parker, 1999; Schmitt, Nason, Whitney, & Pulakos, 1995; Scullen, 1999; Williams & Anderson, 1994; Williams & Brown, 1994).

Method biases are a problem because they are one of the main sources of measurement error. Measurement error threatens the validity of the conclusions about the relationships between measures and is widely recognized to have both a random and a systematic component (cf. Bagozzi & Yi, 1991; Nunnally, 1978; Spector, 1987). Although both types of measurement error are problematic, systematic measurement error is a particularly serious problem because it provides an alternative explanation for the observed relationships between measures of different constructs that is independent of the one hypothesized. Bagozzi and Yi (1991) noted that one of the main sources of systematic measurement error is method variance that may arise from a variety of sources:

> Method variance refers to variance that is attributable to the measurement method rather than to the construct of interest. The term *method* refers to the form of measurement at different levels of abstraction, such as the content of specific items, scale type, response format, and the general context (Fiske, 1982, pp. 81–84). At a more abstract level, method effects might be interpreted in terms of response biases such as halo effects, social desirability, acquiescence, leniency effects, or yea- and nay-saying. (p. 426)

However, regardless of its source, systematic error variance can have a serious confounding influence on empirical results, yielding potentially misleading conclusions (Campbell & Fiske, 1959). For example, let's assume that a researcher is interested in studying a hypothesized relationship between Constructs A and B. Based on theoretical considerations, one would expect that the measures of Construct A would be correlated with measures of Construct B. However, if the measures of Construct A and the measures of Construct B also share common methods, those methods may exert a systematic effect on the observed correlation between the measures. Thus, at least partially, common method biases pose a rival explanation for the correlation observed between the measures.

Within the above context, the purpose of this research is to (a) examine the extent to which method biases influence behavioral research results, (b) identify potential sources of method biases, (c) discuss the cognitive processes through which method biases influence responses to measures, (d) evaluate the many different procedural and statistical techniques that can be used to control method biases, and (e) provide recommendations for how to select appropriate procedural and statistical remedies for different types of research settings. This is important because, to our knowledge, there is no comprehensive discussion of all of these issues available in the literature, and the evidence suggests that many researchers are not effectively controlling for this source of bias.

Extent of the Bias Caused by Common Method Variance

Over the past few decades, a considerable amount of evidence has accumulated regarding the extent to which method variance influences (a) measures used in the field and (b) relationships between these measures. Much of the evidence of the extent to which method variance is present in measures used in behavioral research comes from meta-analyses of multitrait–multimethod

Philip M. Podsakoff and Jeong-Yeon Lee, Department of Management, Indiana University; Scott B. MacKenzie, Department of Marketing, Indiana University; Nathan P. Podsakoff, Department of Management, University of Florida.

Correspondence concerning this article should be addressed to Philip M. Podsakoff, Department of Management, Kelley School of Business, Indiana University, 1309 East Tenth Street, Bloomington, Indiana 47405-1701. E-mail: podsakof@indiana.edu

studies (cf. Bagozzi & Yi, 1990; Cote & Buckley, 1987, 1988; Williams, Cote, & Buckley, 1989). Perhaps the most comprehensive evidence comes from Cote and Buckley (1987), who examined the amount of common method variance present in measures across 70 MTMM studies in the psychology–sociology, marketing, business, and education literatures. They found that approximately one quarter (26.3%) of the variance in a typical research measure might be due to systematic sources of measurement error like common method biases. However, they also found that the amount of variance attributable to method biases varied considerably by discipline and by the type of construct being investigated. For example, Cote and Buckley (1987) found that, on average, method variance was lowest in the field of marketing (15.8%) and highest in the field of education (30.5%). They also found that typical job performance measures contained an average of 22.5% method variance, whereas attitude measures contain an average of 40.7%. A similar pattern of findings emerges from Williams et al.'s (1989) study of just the applied psychology literature.

In addition to these estimates of the extent to which method variance is present in typical measures, there is also a growing body of research examining the extent to which method variance influences relationships between measures (cf. Fuller, Patterson, Hester, & Stringer, 1996; Gerstner & Day, 1997; Lowe, Kroeck, & Sivasubramaniam, 1996; Podsakoff, MacKenzie, Paine, & Bachrach, 2000; Wagner & Gooding, 1987). These studies contrasted the strength of the relationship between two variables when common method variance was controlled versus when it was not. They found that, on average, the amount of variance accounted for when common method variance was present was approximately 35% versus approximately 11% when it was not present. Thus, there is a considerable amount of evidence that common method variance can have a substantial effect on observed relationships between measures of different constructs. However, it is important to recognize that the findings suggest that the magnitude of the bias produced by these method factors varies across research contexts (cf. Cote & Buckley, 1987; Crampton & Wagner, 1994).

Not only can the strength of the bias vary but so can the direction of its effect. Method variance can either inflate or deflate observed relationships between constructs, thus leading to both Type I and Type II errors. This point is illustrated in Table 1, which uses Cote and Buckley's (1987) estimates of the average amount of trait variance, the average amount of method variance, and the average method intercorrelations and inserts them into the equation below to calculate the impact of common method variance on the observed correlation between measures of different types of constructs (e.g., attitude, personality, aptitude):

$$R_{x,y} = (\text{true } R_{ti,tj}\sqrt{t_x}\sqrt{t_y}) + (\text{true } R_{mk,ml}\sqrt{m_x}\sqrt{m_y}), \qquad (1)$$

where true $R_{ti,tj}$ is the average correlation between trait i and trait j, t_x is the percent of trait variance in measure x, t_y is the percent of trait variance in measure y, true $R_{mk,ml}$ is the average correlation between method k and method l, m_x is the percent of method variance in measure x, and m_y is the percent of method variance in measure y.

For example, the correlation .52 in the second row of the first column of Table 1 was calculated by multiplying the true correlation (1.00) times the square root of Cote and Buckley's (1987) estimate of the percent of trait variance typically found in attitude measures ($\sqrt{.298}$) times the square root of their estimate of the percent of trait variance typically found in personality measures ($\sqrt{.391}$) plus the average of their estimates of the typical correlation between methods for attitude (.556) and personality (.546) constructs multiplied by the square root of their estimate of the percent of method variance typically found in attitude measures ($\sqrt{.407}$) times the square root of their estimate of the percent of method variance typically found in personality measures ($\sqrt{.247}$).

Table 1
Relationship Between True and Observed Correlation for Average Measures by Type of Construct

Type of Constructs	True $R_{ti,tj}$ correlation ($R_{ti,tj}^2$)				
	1.00 (1.00)	.50 (.25)	.30 (.09)	.10 (.01)	.00 (.00)
Attitude–attitude	.52 (.27)	.38 (.14)	.32 (.10)	.26 (.07)	.23 (.05)
Attitude–personality	.52 (.27)	.35 (.12)	.28 (.08)	.21 (.04)	.17 (.03)
Attitude–aptitude	.52 (.27)	.35 (.12)	.28 (.08)	.21 (.04)	.18 (.03)
Attitude–job performance and satisfaction	.51 (.26)	.32 (.10)	.25 (.06)	.17 (.03)	.13 (.02)
Personality–personality	.53 (.28)	.33 (.11)	.25 (.06)	.17 (.03)	.13 (.02)
Personality–aptitude	.53 (.28)	.34 (.12)	.26 (.07)	.18 (.03)	.14 (.02)
Personality–job performance and satisfaction	.53 (.28)	.32 (.10)	.23 (.05)	.15 (.02)	.10 (.01)
Aptitude–aptitude	.54 (.29)	.34 (.12)	.26 (.07)	.18 (.03)	.14 (.02)
Aptitude–job performance and satisfaction	.54 (.29)	.32 (.10)	.24 (.06)	.15 (.02)	.11 (.01)
Job performance and satisfaction–job performance and satisfaction	.54 (.29)	.31 (.09)	.21 (.04)	.12 (.01)	.07 (.00)

Note. Values within the table are the observed correlations $R_{x,y}$ (and squared correlations $R_{x,y}^2$) calculated using Cote and Buckley's (1988) formula shown in Equation 1 of the text. For the calculations it is assumed that (a) the trait variance is the same as that reported by Cote and Buckley (1987) for each type of construct (e.g., attitude measures = .298, personality measures = .391, aptitude measures = .395, and job performance and satisfaction measures = .465), (b) the method variance is the same as that reported by Cote and Buckley (1987) for each type of construct (e.g., attitude measures = .407, personality measures = .247, aptitude measures = .251, and job performance and satisfaction measures = .225), and (c) the correlation between the methods is the average of the method correlations reported by Cote and Buckley (1987) for each of the constructs (e.g., method correlations between attitude–attitude constructs = .556, personality–attitude constructs = .551, personality–personality constructs = .546, aptitude–attitude constructs = .564, aptitude–personality constructs = .559, aptitude–aptitude constructs = .572, job performance and satisfaction–attitude constructs = .442, job performance and satisfaction–personality constructs = .437, job performance and satisfaction–aptitude constructs = .450, and job performance and satisfaction–job performance and satisfaction constructs = .328). These calculations ignore potential Trait × Method interactions.

There are several important conclusions that can be drawn from Table 1. For example, the entry in the first column of the first row indicates that even though two attitude constructs are perfectly correlated, the observed correlation between their measures is only .52 because of measurement error. Similarly, the entry in the last column of the first row indicates that even though two attitude constructs are completely uncorrelated, the observed correlation between their measures is .23 because of random and systematic measurement error. Both of these numbers are troubling but for different reasons. The entries in the entire first column are troubling because they show that even though two traits are perfectly correlated, typical levels of measurement error cut the observed correlation between their measures in half and the variance explained by 70%. The last column of entries is troubling because it shows that even when two constructs are completely uncorrelated, measurement error causes the observed correlation between their measures to be greater than zero. Indeed, some of these numbers are not very different from the effect sizes reported in the behavioral literature. In view of this, it is disturbing that most studies ignore measurement error entirely and that even many of the ones that do try to take random measurement error into account ignore systematic measurement error. Thus, measurement error can inflate or deflate the observed correlation between the measures, depending on the correlation between the methods. Indeed, as noted by Cote and Buckley (1988), method effects inflate the observed relationship when the correlation between the methods is higher than the observed correlation between the measures with method effects removed and deflate the relationship when the correlation between the methods is lower than the observed correlation between the measures with method effects removed.

Potential Sources of Common Method Biases

Because common method biases can have potentially serious effects on research findings, it is important to understand their sources and when they are especially likely to be a problem. Therefore, in the next sections of the article, we identify several of the most likely causes of method bias and the research settings in which they are likely to pose particular problems. As shown in Table 2, some sources of common method biases result from the fact that the predictor and criterion variables are obtained from the same source or rater, whereas others are produced by the measurement items themselves, the context of the items within the measurement instrument, and/or the context in which the measures are obtained.

Method Effects Produced by a Common Source or Rater

Some methods effects result from the fact that the respondent providing the measure of the predictor and criterion variable is the same person. This type of self-report bias may be said to result from any artifactual covariance between the predictor and criterion variable produced by the fact that the respondent providing the measure of these variables is the same.

Consistency motif. There is a substantial amount of theory (cf. Heider, 1958; Osgood & Tannenbaum, 1955) and research (cf. McGuire, 1966) suggesting that people try to maintain consistency between their cognitions and attitudes. Thus, it should not be surprising that people responding to questions posed by researchers would have a desire to appear consistent and rational in their responses and might search for similarities in the questions asked of them—thereby producing relationships that would not otherwise exist at the same level in real-life settings. This tendency of respondents to try to maintain consistency in their responses to similar questions or to organize information in consistent ways is called the *consistency motif* (Johns, 1994; Podsakoff & Organ, 1986; Schmitt, 1994) or the *consistency effect* (Salancik & Pfeffer, 1977) and is likely to be particularly problematic in those situations in which respondents are asked to provide retrospective accounts of their attitudes, perceptions, and/or behaviors.

Implicit theories and illusory correlations. Related to the notion of the consistency motif as a potential source of common method variance are *illusory correlations* (cf. Berman & Kenny, 1976; Chapman & Chapman, 1967, 1969; Smither, Collins, & Buda, 1989), and *implicit theories* (cf. Lord, Binning, Rush, & Thomas, 1978; Phillips & Lord, 1986; Staw, 1975). Berman and Kenny (1976) have indicated that illusory correlations result from the fact that "raters often appear to possess assumptions concerning the co-occurrence of rated items, and these assumptions may introduce systematic distortions when correlations are derived from the ratings" (p. 264); Smither et al. (1989) have noted that these "illusory correlations may serve as the basis of job schema or implicit theories held by raters and thereby affect attention to and encoding of ratee behaviors as well as later recall" (p. 599). This suggests that correlations derived from ratees' responses are composed of not only true relationships but also artifactual covariation based on ratees' implicit theories.

Indeed, there is a substantial amount of evidence that implicit theories do have an effect on respondents' ratings in a variety of different domains, including ratings of leader behavior (e.g., Eden & Leviatin, 1975; Lord et al., 1978; Phillips & Lord, 1986), attributions of the causes of group performance (cf. Guzzo, Wagner, Maguire, Herr, & Hawley, 1986; Staw, 1975), and perceptions about the relationship between employee satisfaction and performance (Smither et al., 1989). Taken together, these findings indicate that the relationships researchers observe between predictor and criterion variables on a questionnaire may not only reflect the actual covariation that exists between these events but may also be the result of the implicit theories that respondents have regarding the relationship between these events.

Social desirability. According to Crowne and Marlowe (1964), *social desirability* "refers to the need for social approval and acceptance and the belief that it can be attained by means of culturally acceptable and appropriate behaviors" (p. 109). It is generally viewed as the tendency on the part of individuals to present themselves in a favorable light, regardless of their true feelings about an issue or topic. This tendency is problematic, not only because of its potential to bias the answers of respondents (i.e., to change the mean levels of the response) but also because it may mask the true relationships between two or more variables (Ganster, Hennessey & Luthans, 1983). Ganster et al. (1983) have noted that social desirability can produce spurious relationships, serve as a suppressor variable that hides the true relationship between variables, or serve as a moderator variable that influences the nature of the relationships between the variables.

Leniency biases. Guilford (1954, p. 278) has defined *leniency biases* as the tendency for raters "to rate those whom they know well, or whom they are ego involved, higher than they should." Research on this form of bias (Schriesheim, Kinicki, &

Table 2
Summary of Potential Sources of Common Method Biases

Potential cause	Definition
Common rater effects	Refer to any artifactual covariance between the predictor and criterion variable produced by the fact that the respondent providing the measure of these variables is the same.
Consistency motif	Refers to the propensity for respondents to try to maintain consistency in their responses to questions.
Implicit theories (and illusory correlations)	Refer to respondents' beliefs about the covariation among particular traits, behaviors, and/or outcomes.
Social desirability	Refers to the tendency of some people to respond to items more as a result of their social acceptability than their true feelings.
Leniency biases	Refer to the propensity for respondents to attribute socially desirable traits, attitudes, and/or behaviors to someone they know and like than to someone they dislike.
Acquiescence biases (yea-saying and nay-saying)	Refer to the propensity for respondents to agree (or disagree) with questionnaire items independent of their content.
Mood state (positive or negative affectivity; positive or negative emotionality)	Refers to the propensity of respondents to view themselves and the world around them in generally negative terms (negative affectivity) or the propensity of respondents to view themselves and the world around them in generally positive terms (positive affectivity).
Transient mood state	Refers to the impact of relatively recent mood-inducing events to influence the manner in which respondents view themselves and the world around them.
Item characteristic effects	Refer to any artifactual covariance that is caused by the influence or interpretation that a respondent might ascribe to an item solely because of specific properties or characteristics the item possesses.
Item social desirability	Refers to the fact that items may be written in such a way as to reflect more socially desirable attitudes, behaviors, or perceptions.
Item demand characteristics	Refer to the fact that items may convey hidden cues as to how to respond to them.
Item ambiguity	Refers to the fact that items that are ambiguous allow respondents to respond to them systematically using their own heuristic or respond to them randomly.
Common scale formats	Refer to artifactual covariation produced by the use of the same scale format (e.g., Likert scales, semantic differential scales, "faces" scales) on a questionnaire.
Common scale anchors	Refer to the repeated use of the same anchor points (e.g., *extremely*, *always*, *never*) on a questionnaire.
Positive and negative item wording	Refers to the fact that the use of positively (negatively) worded items may produce artifactual relationships on the questionnaire.
Item context effects	Refer to any influence or interpretation that a respondent might ascribe to an item solely because of its relation to the other items making up an instrument (Wainer & Kiely, 1987).
Item priming effects	Refer to the fact that the positioning of the predictor (or criterion) variable on the questionnaire can make that variable more salient to the respondent and imply a causal relationship with other variables.
Item embeddedness	Refers to the fact that neutral items embedded in the context of either positively or negatively worded items will take on the evaluative properties of those items.
Context-induced mood	Refers to when the first question (or set of questions) encountered on the questionnaire induces a mood for responding to the remainder of the questionnaire.
Scale length	Refers to the fact that if scales have fewer items, responses to previous items are more likely to be accessible in short-term memory and to be recalled when responding to other items.
Intermixing (or grouping) of items or constructs on the questionnaire	Refers to the fact that items from different constructs that are grouped together may decrease intraconstruct correlations and increase interconstruct correlations.
Measurement context effects	Refer to any artifactual covariation produced from the context in which the measures are obtained.
Predictor and criterion variables measured at the same point in time	Refers to the fact that measures of different constructs measured at the same point in time may produce artifactual covariance independent of the content of the constructs themselves.
Predictor and criterion variables measured in the same location	Refers to the fact that measures of different constructs measured in the same location may produce artifactual covariance independent of the content of the constructs themselves.
Predictor and criterion variables measured using the same medium	Refers to the fact that measures of different constructs measured with the same medium may produce artifactual covariance independent of the content of the constructs themselves.

Schriesheim, 1979) has shown that it produces spurious correlations between leader-consideration behavior and employee satisfaction and perceptions of group productivity, drive, and cohesiveness but not between leader initiating structure behavior and these same criterion variables. This suggests that the consideration scale is not socially neutral and that leniency biases tend to influence the relationships obtained between this scale and employee attitudes and perceptions. One might also expect leniency biases to produce spurious correlations in other studies that examine the relationship between respondents' ratings of liked (or disliked) others and the respondents' ratings of the performance, attitudes, and perceptions of others.

Acquiescence (yea-saying or nay-saying). Winkler, Kanouse, and Ware (1982, p. 555) have defined *acquiescence response set* as the "tendency to agree with attitude statements regardless of content" and have noted that this response set is problematic "because it heightens the correlations among items that are worded similarly, even when they are not conceptually related." Although Winkler et al. (1982) focused specific attention on the effects of acquiescence on scale development processes, it is easy to see how this form of bias might also cause spurious relationships between two or more constructs. Thus, acquiescence may also be a potential cause of artifactual variance in the relationships between two or

more variables, other than the true variance between these variables.

Positive and negative affectivity. Waston and Clark (1984) defined *negative affectivity* as a mood-dispositional dimension that reflects pervasive individual differences in negative emotionality and self-concept and *positive affectivity* as reflecting pervasive individual differences in positive emotionality and self-concept. On the basis of their extensive review of the literature, Watson and Clark concluded that people who express high negative affectivity view themselves and a variety of aspects of the world around them in generally negative terms. Burke, Brief, and George (1993) drew similar conclusions regarding the potential impact of positive affectivity. They noted that,

> Self-reports of negative features of the work situation and negative affective reactions may both be influenced by negative affectivity, whereas self-reports of positive aspects of the work situation and positive affective reactions may both be influenced by positive affectivity. (Burke et al., 1993, p. 410)

If these dispositions influence respondents' ratings on self-report questionnaires, it is possible that negative (positive) affectivity could account for systematic variance in the relationships obtained between two or more variables that is different from the actual (true) score variance that exists between these variables. Indeed, Brief, Burke, George, Robinson, and Webster (1988) reported that negative affectivity inflated the relationships obtained between expressions of employee stress and their expressions of job and life satisfaction, depression, and the amount of negative affect experienced at work, and Williams and Anderson (1994) reported that their structural equation models specifying the relationships between leader contingent reward behavior and job satisfaction and commitment fit significantly better when their measure of positive affectivity was included in the model than when it was excluded from the model. In contrast, Chen and Spector (1991) and Jex and Spector (1996) found little support for the influence of negative affectivity on the relationships between self-reported job stress and job strain variables. Although these contradictory findings have led to a fairly lively debate (cf. Bagozzi & Yi, 1990; Spector, 1987; Williams et al., 1989), taken as a whole, they appear to indicate that positive and negative affectivity may influence the relationships between variables in organizational research.

Transient mood state. Positive and negative affectivity are generally considered to be fairly enduring trait characteristics of the individual that may influence their responses to questionnaires. However, it is also possible that the *transient mood states* of respondents produced from any of a number of events (e.g., interaction with a disgruntled customer, receiving a compliment from a co-worker or boss, receiving word of a promotion, death of a close friend or family member, a bad day at the office, concerns about downsizing) may also produce artifactual covariance in self-report measures because the person responds to questions about both the predictor and criterion variable while in a particular mood.

Method Effects Produced by Item Characteristics

In addition to the bias that may be produced by obtaining measures from the same source, it is also possible for the manner in which items are presented to respondents to produce artifactual covariance in the observed relationships. Cronbach (Cronbach, 1946, 1950) was probably the first to recognize the possibility that, in addition to its content, an item's form may also influence the scores obtained on a measure:

> A psychological test or educational test is constructed by choosing items of the desired content, and refining them by empirical techniques. The assumption is generally made, and validated as well as possible, that what the test measures is determined by the content of the items. Yet the final score of the person on any test is a composite of effects resulting from the content of the item and effects resulting from the form of the item used. A test supposedly measuring one variable may also be measuring another trait which would not influence the score if another type of item were used. (Cronbach, 1946, pp. 475–476)

Although Cronbach's (1946, 1950) discussion of response sets tended to confound characteristics of the items of measurement with personal tendencies on the part of respondents exposed to those items, our focus in this section is on the potential effects that item characteristics have on common method variance.

Item social desirability (or item demand characteristics). Thomas and Kilmann (1975) and Nederhof (1985) have noted that, in addition to the fact that social desirability may be viewed as a tendency for respondents to behave in a culturally acceptable and appropriate manner, it may also be viewed as a property of the items in a questionnaire. As such, items or constructs on a questionnaire that possess more (as opposed to less) social desirability may be observed to relate more (or less) to each other as much because of their social desirability as they do because of the underlying constructs that they are intended to measure. For example, Thomas and Kilmann (1975) reported that respondents' self-reported ratings of their use of five different conflict-handling modes of behavior correlated strongly with the rated social desirability of these modes of behavior. Thus, social desirability at the item (and/or construct) level is also a potential cause of artifactual variance in questionnaire research.

Item complexity and/or ambiguity. Although researchers are encouraged to develop items that are as clear, concise, and specific as possible to measure the constructs they are interested in (cf. Peterson, 2000; Spector, 1992), it is not uncommon for some items to be fairly complex or ambiguous. Undoubtedly, some of the complexity that exists in questionnaire measures results from the fact that some constructs are fairly complex or abstract in nature. However, in other cases, *item complexity or ambiguity* may result from the use of double-barreled questions (cf. Hinkin, 1995), words with multiple meanings (Peterson, 2000), technical jargon or colloquialisms (Spector, 1992), or unfamiliar or infrequently used words (Peterson, 2000). The problem with ambiguous items is that they often require respondents to develop their own idiosyncratic meanings for them. This may either increase random responding or increase the probability that respondents' own systematic response tendencies (e.g., implicit theories, affectivity, central tendency and leniency biases) may come into play. For example, Gioia and Sims (1985) reported that implicit leadership theories are more likely to influence respondents' ratings when the leader behaviors being rated are less behaviorally specific than when the leader behaviors being rated are more behaviorally specific. Thus, in addition to item content, the level of item

ambiguity and complexity may also influence the relationships obtained between the variables of interest in a study.

Scale format and scale anchors. It is not uncommon for researchers to measure different constructs with similar *scale formats* (e.g., Likert scales, semantic differential scales, "faces" scales), using similar *scale anchors* or values ("extremely" vs. "somewhat," "always" vs. "never," and "strongly agree" vs. "strongly disagree"). Although it may be argued that the use of similar scale formats and anchors makes it easier for the respondents to complete the questionnaire because it provides a standardized format and therefore requires less cognitive processing, this may also increase the possibility that some of the covariation observed among the constructs examined may be the result of the consistency in the scale properties rather than the content of the items. For example, it is well known in the survey research literature (Tourangeau, Rips, & Rasinski, 2000) that scale format and anchors systematically influence responses.

Negatively worded (reverse-coded) items. Some researchers have attempted to reduce the potential effects of response pattern biases by incorporating negatively worded or reverse-coded items on their questionnaires (cf. Hinkin, 1995; Idaszak & Drasgow, 1987). The basic logic here is that *reverse-coded items* are like cognitive "speed bumps" that require respondents to engage in more controlled, as opposed to automatic, cognitive processing. Unfortunately, research has shown that reverse-coded items may produce artifactual response factors consisting exclusively of negatively worded items (Harvey, Billings, & Nilan, 1985) that may disappear after the reverse-coded items are rewritten in a positive manner (Idaszak & Drasgow, 1987). Schmitt and Stults (1986) have argued that the effects of *negatively worded items* may occur because once respondents establish a pattern of responding to a questionnaire, they may fail to attend to the positive–negative wording of the items. In addition, they have shown that factors representing negatively worded items may occur in cases where as few as 10% of the respondents fail to recognize that some items are reverse coded. Thus, negatively worded items may be a source of method bias.

Method Effects Produced by Item Context

Common method biases may also result from the context in which the items on a questionnaire are placed. Wainer and Keily (1987) have suggested that *item context effects* "refer to any influence or interpretation that a subject might ascribe to an item solely because of its relation to the other items making up an instrument" (p. 187). In this section, we examine several potential types of item context effects.

Item priming effects. Salancik and Pfeffer (1977) and Salancik (1984) have noted that asking questions about particular features of the work environment may make other work aspects more salient to respondents than these work aspects would have been if the questions had not been asked in the first place. They referred to this increased salience as a "priming" effect and described it in the context of the need–satisfaction models by noting that

> If a person is asked to describe his job in terms that are of interest to the investigator, he can do so. But if the individual is then asked how he feels about the job, he has few options but to respond using the information the investigator has made salient. The correlation between job characteristics and attitudes from such a study is not only unremarkable, but provides little information about the need–satisfaction model. (Salancik & Pfeffer, 1977, p. 451)

Although the specific conditions identified by Salancik (1984) as necessary to produce priming effects have been the subject of some debate (cf. Salancik, 1982, 1984; Stone, 1984; Stone & Gueutal, 1984), there is some evidence that priming effects may occur in some instances (cf. Salancik, 1982). Thus, it is possible for such effects to produce artifactual covariation among variables under some conditions.

Item embeddedness. Harrison and McLaughlin (1993) have argued that neutral items embedded in the context of either positively or negatively worded items take on the evaluative properties of those items and that this process may subsequently influence the observed covariation among these items. More specifically, they noted that

> item context can influence a respondent's interpretation of a question, retrieval of information from memory, judgment about that retrieved information, and the selection of an appropriate item response. An item's context can cause cognitive carryover effects by influencing any number of these processing stages. Carryover occurs when the interpretation, retrieval, judgment, or response associated with prior items provides a respondent with an easily accessible cognitive structure or schema, by bringing the cognitive structure into short-term memory, into a temporary *workspace*, or to the top of the storage bin of relevant information . . . A respondent then uses the easily accessible set of cognitions to answer subsequent items. Cognitive carryover can produce spurious response consistency in attitude surveys in the same way it can produce illusory halo error in performance ratings. (Harrison & McLaughlin, 1993, p. 131)

On the basis of these arguments regarding cognitive carryover effects, Harrison and McLaughlin (1993) predicted and found that evaluatively neutral items placed in blocks of positive (or negative) evaluative items were rated in a manner similar to the items they were embedded within. Another example of item embeddedness effects is the "chameleon effect" noted by Marsh and Yeung (1999). They found that answers to general self-esteem questions (e.g., "I feel good about myself") reflected the nature of the surrounding questions. More specifically, they demonstrated that the content-free esteem item could take on qualitatively different meanings, depending on the context in which it appears. Thus, it appears that the context in which neutral items are presented may influence the manner in which these items are rated.

Context-induced mood. Earlier, we noted that respondents' moods (whether stable or based on transient events) may influence their responses to questionnaire items, independent of the content of the items themselves. One factor that might produce transient mood states on the part of respondents is the manner in which the items on the questionnaire are worded (cf. Peterson, 2000). For example, it is possible that the wording of the first set of items on a questionnaire induces a mood on the part of respondents that influences the manner in which they respond to the remaining items on the questionnaire. Thus, items on a questionnaire that raise respondents' suspicions about the researcher's intent or integrity or items that insult the respondent, because they relate to ethnic, racial, or gender stereotypes, might predispose the respondent to complete the questionnaire with a negative mood state. It is possible for these context-induced moods to produce artifactual covariation among the constructs on the questionnaire.

Scale length. Harrison, McLaughlin, and Coalter (1996) have noted that scales that contain fewer items increase respondents' accessibility to answers to previous scales, thereby increasing the likelihood that these previous responses influence answers to current scales. Their logic is that shorter scales minimize the decay of previous responses in short-term memory, thereby enhancing the observed relationships between scale items that are similar in content. Therefore, although scales that are short in length have some advantages in that they may reduce some forms of bias that are produced by respondent fatigue and carelessness (cf. Hinkin, 1995), they may actually enhance other forms of bias because they increase the possibility that responses to previous items on the questionnaire will influence responses to current items.

Intermixing items of different constructs on the questionnaire. It is not uncommon for researchers to intermix items from different constructs on the same questionnaire. Indeed, Kline et al. (2000) recommended this practice to reduce common method variance. However, if the constructs on the questionnaire are similar (like job characteristics and job satisfaction), one possible outcome of this practice is that it may increase the interconstruct correlations at the same time it decreases the intraconstruct correlations. This would appear to suggest that intermixing items on a questionnaire would produce artifactual covariation among the constructs.

However, the issue is probably more complex than it appears to be on the surface. For example, at the same time that the intermixed items may increase the interconstruct correlations because respondents have a more difficult time distinguishing between the constructs, the reliability in the scales may be reduced because the respondents have a more difficult time also seeing the similarity in the items measuring the same construct. However, reducing the reliability of the scales should have the effect of reducing (rather than increasing) the covariation among the constructs. Thus, it is difficult to tell how the countervailing effects of increasing interconstruct correlations at the same time as decreasing intraconstruct correlations affects method variance. Therefore, it appears that more attention needs to be directed at this issue before we can really make definitive statements regarding the effects of mixed versus grouped items on method variance.

Method Effects Produced by Measurement Context

A final factor that may influence the artifactual covariation observed between constructs is the broader research context in which the measures are obtained. Chief among these contextual influences are the time, location, and media used to measure the constructs.

Time and location of measurement. Measures of predictor and criterion variables may be assessed concurrently or at different times and places. To the extent that measures are taken at the same time in the same place, they may share systematic covariation because this common measurement context may (a) increase the likelihood that responses to measures of the predictor and criterion variables will co-exist in short-term memory, (b) provide contextual cues for retrieval of information from long-term memory, and (c) facilitate the use of implicit theories when they exist.

Use of common medium to obtain measurement. One final contextual factor that may produce artifactual covariation among the predictor and criterion variables is the medium used to obtain the responses. For example, interviewer characteristics, expectations, and verbal idiosyncrasies are well recognized in the survey response literature as potential sources of method biases (cf. Bouchard, 1976; Collins, 1970; Shapiro, 1970). Similarly, research (cf. Martin & Nagao, 1989; Richman, Kiesler, Weisband, & Drasgow, 1999) has shown that face-to-face interviews tend to induce more socially desirable responding and lower accuracy than computer-administered questionnaires or paper-and-pencil questionnaires. Thus, the medium used to gather data may be a source of common method variance.

Summary of the Sources of Common Method Variance

In summary, common method biases arise from having a common rater, a common measurement context, a common item context, or from the characteristics of the items themselves. Obviously, in any given study, it is possible for several of these factors to be operative. Therefore, it is important to carefully evaluate the conditions under which the data are obtained to assess the extent to which method biases may be a problem. Method biases are likely to be particularly powerful in studies in which the data for both the predictor and criterion variable are obtained from the same person in the same measurement context using the same item context and similar item characteristics. These conditions are often present in behavioral research. For example, Sackett and Larson (1990) reviewed every research study appearing in *Journal of Applied Psychology*, *Organizational Behavior and Human Decision Processes*, and *Personnel Psychology* in 1977, 1982, and 1987 and found that 51% (296 out of 577) of all the studies used some kind of self-report measure as either the primary or sole type of data gathered and were therefore subject to common rater biases. They also found that 39% (222 out of 577) used a questionnaire or interview methodology wherein all of the data were collected in the same measurement context.

Processes Through Which Method Biases Influence Respondent Behavior

Once the method biases that are likely to be present in a particular situation have been identified, the next step is to develop procedures to minimize their impact. However, to do this, one must understand how these biases affect the response process. Although there are many different models of how people generate responses to questions (cf. Cannell, Miller, & Oksenberg, 1981; Strack & Martin, 1987; Thurstone, 1927; Tourangeau et al., 2000), there are several similarities with respect to the fundamental stages they include. The first two columns of Table 3 show the most commonly identified stages of the response process: comprehension, retrieval, judgment, response selection, and response reporting (cf. Strack & Martin, 1987; Sudman, Bradburn, & Schwarz, 1996; Tourangeau et al., 2000). In the comprehension stage, respondents attend to the questions and instructions they receive and try to understand what the question is asking. The retrieval stage involves generating a retrieval strategy and a set of cues that can be used to recall relevant information from long-term memory. However, because retrieval does not always yield an explicit answer to the questions being asked, the next step is for the respondent to assess the completeness and accuracy of their memories, draw inferences that fill in gaps in what is recalled, and integrate the material retrieved into a single overall judgment of

Table 3
How Common Method Biases Influence the Question Response Process

Stages of the response process	Activities involved in each stage	Potential method biases
Comprehension	Attend to questions and instructions, represent logical form of question, identify information sought, and link key terms to relevant concepts	Item ambiguity
Retrieval	Generate retrieval strategy and cues, retrieve specific and generic memories, and fill in missing details	Measurement context, question context, item embeddedness, item intermixing, scale size, priming effects, transient mood states, and item social desirability
Judgment	Assess completeness and accuracy of memories, draw inferences based on accessibility, inferences that fill in gaps of what is recalled, integrate material retrieved, and make estimate based on partial retrieval	Consistency motif (when it is an attempt to increase accuracy in the face of uncertainty), implicit theories, priming effects, item demand characteristics, and item context-induced mood states
Response selection	Map judgment onto response category	Common scale anchors and formats and item context-induced anchoring effects
Response reporting	Editing response for consistency, acceptability, or other criteria	Consistency motif (when it is an attempt to appear rational), leniency bias, acquiescence bias, demand characteristics, and social desirability

Note. The first two columns of this table are adapted from *The Psychology of Survey Response*, by R. Tourangeau, L. J. Rips, and K. Rasinski, 2000, Cambridge, England: Cambridge University Press. Copyright 2000 by Cambridge University Press. Adapted with permission.

how to respond. Once respondents have identified what they think that the answer is, their next task is to decide how their answer maps onto the appropriate scale or response option provided to them. Following this, respondents either record their judgment on the appropriate point on the scale or edit their responses for consistency, acceptability, desirability, or other criteria. In describing these stages in the response process, we are not suggesting that the response process is always highly conscious and deliberative. Indeed, anyone who has observed people filling out questionnaires would agree that all of these stages of this process might happen very quickly in a more or less automatic manner.

In the third column of Table 3, we have identified the specific method biases that are likely to have the biggest effects at each stage of the response process. Several researchers (cf. Fowler, 1992; Torangeau et al., 2000) have noted that item ambiguity is likely to be the biggest problem in the comprehension stage because the more ambiguous the question is, the more difficult it is for respondents to understand what they are being asked and how to link the question to relevant concepts and information in memory. A review of the literature by Sudman et al. (1996) suggested that when faced with an ambiguous question, respondents often refer to the surrounding questions to infer the meaning of the ambiguous one. This causes the answers to the surrounding questions to be systematically related to the answer to the ambiguous question. Alternatively, when a question is highly ambiguous, respondents may respond either systematically by using some heuristic (e.g., some people may respond neutrally, whereas others may agree or disagree) or randomly without using any heuristic at all. To the extent that people rely on heuristics when responding to ambiguous questions, it could increase common method variance between an ambiguous predictor and an ambiguous criterion variable.

There are several common method biases that may affect the retrieval stage of the response process. These biases influence this stage by providing common cues that influence what is retrieved from memory and thereby influence the correlation between the measures of the predictor and criterion variables. For example, measuring both the predictor and criterion variable in the same measurement context (in terms of time, location, position in the questionnaire) can provide common contextual cues that influence the retrieval of information from memory and the correlation between the two measures (cf. Sudman et al., 1996). Mood is another method bias affecting the retrieval stage of the response process. A great deal of research indicates that transient mood states and/or context-induced mood states influence the contents of what is recalled from memory (cf. Blaney, 1986; Bower, 1981; Isen & Baron, 1991; Parrott & Sabini, 1990). Some research has shown mood-congruency effects on recall (Bower, 1981; Isen & Baron, 1991), whereby an individual's current mood state automatically primes similarly valenced material stored in memory, and some has shown mood-incongruency effects (Parrott & Sabini, 1990) that are typically attributed to motivational factors. However, all of this research strongly supports the notion that mood affects recall in a systematic manner. When this mood is induced by the measurement context and/or when it is present when a person responds to questions about both the predictor and criterion variables, its biasing effects on retrieval are likely to be a source of common method bias.

As indicated in Table 3, there are also several method biases that affect the judgment stage of the response process. Some of these affect the process of drawing inferences that fill in gaps in what is recalled. For example, implicit theories may be used to fill in gaps in what is recalled or to infer missing details from memory on the basis of what typically happens. Similarly, item demand characteristics may prompt people who are uncertain about how to respond on the basis of the cues present in the question itself. Likewise, people may rely on a consistency motif to fill in the missing information when faced with gaps in what is recalled or uncertainty regarding the accuracy of the information recalled from memory (e.g., "Since I remember doing *X*, I probably also did *Y*").

Other method biases may affect the judgment stage by influencing the process of making estimates based on the partial retrieval of information. For example, priming effects may influence

the judgment stage of the response process because answering initial questions brings information into short-term memory that remains accessible when responding to later questions (Judd, Drake, Downing, & Krosnick, 1991; Salancik, 1984; Torangeau, Rasinski, & D'Andrade, 1991). The same can be said for mood produced by the item context. In general, when questions change a respondent's current mood by bringing positive or negative material to mind, it is likely to affect subsequent judgments even if the target of judgment is completely unrelated (Sudman et al., 1996). Finally, in addition to item-context-induced mood, the item context may also produce other experiences that affect subsequent judgments. For example, respondents may find it easy or difficult to answer specific questions, and this subjective experience alone may be used as a heuristic for making subsequent judgments (cf. Sudman et al., 1996).

In the response selection stage, people attempt to map their judgments onto response categories provided by the questions. At this stage, one of the most important method biases is likely to be commonalities in the scale anchors and formats (Tourangeau et al., 2000). For example, some people may be hesitant to say "never" or "always" and therefore select a response option that is less extreme when confronted with a scale anchored with endpoints of "always" and/or "never." When the predictor and criterion variables both share these endpoints, this pattern of responding may artificially enhance the correlation between them. Another method bias operating at this stage occurs when preceding questions influence how respondents use the response scales provided to them. Previous research (cf. Sudman et al., 1996) suggests that when multiple judgments are made by a respondent using the same scale, respondents use their initial ratings to anchor the scale and thereby influence the scaling of their subsequent judgments. In this way, answers to a question may be influenced by the items preceding it on the questionnaire, thus influencing the covariation between the items.

The final stage in the response process involves the editing of the responses for consistency, acceptability, or other criteria. Many of the method biases identified in the literature are involved at this stage. For example, social desirability biases result from the tendency of some people to respond in a socially acceptable manner, even if their true feelings are different from their responses. Consistency biases often reflect the propensity for respondents to try to appear consistent or rational in their responses to questions. Leniency biases reflect the propensity for respondents to rate those that they know well higher than they should. Finally, acquiescence (yea-saying or nay-saying) biases reflect the propensity for respondents to agree (or disagree) with questionnaire items independent of their content. All of these are examples of how people edit their responses prior to reporting them.

Techniques for Controlling Common Method Biases

The previous section identified how different method biases influence the response process. This knowledge can be used to develop procedures to control their effects. Therefore, in this section, we discuss the various ways to control for common method variance and some of the advantages and disadvantages with each of these techniques. Generally speaking, the two primary ways to control for method biases are through (a) the design of the study's procedures and/or (b) statistical controls.

Procedural Remedies

The key to controlling method variance through procedural remedies is to identify what the measures of the predictor and criterion variables have in common and eliminate or minimize it through the design of the study. The connection between the predictor and criterion variable may come from (a) the respondent, (b) contextual cues present in the measurement environment or within the questionnaire itself, and/or (c) the specific wording and format of the questions.

Obtain measures of the predictor and criterion variables from different sources. Because one of the major causes of common method variance is obtaining the measures of both predictor and criterion variables from the same rater or source, one way of controlling for it is to collect the measures of these variables from different sources. For example, those researchers interested in the effects of leader behaviors on employee performance can obtain the measures of leader behavior from the subordinates and the measures of the subordinate's performance from the leader. Similarly, those researchers interested in research on the relationship between organizational culture and organizational performance can obtain the cultural measures from key informants and the measures of organizational performance from archival sources. The advantage of this procedure is that it makes it impossible for the mind set of the source or rater to bias the observed relationship between the predictor and criterion variable, thus eliminating the effects of consistency motifs, implicit theories, social desirability tendencies, dispositional and transient mood states, and any tendencies on the part of the rater to acquiesce or respond in a lenient manner.

Despite the obvious advantages of this approach, it is not feasible to use in all cases. For example, researchers examining the relationships between two or more employee job attitudes cannot obtain measures of these constructs from alternative sources. Similarly, it may not be possible to obtain archival data or to obtain archival data that adequately represent one of the constructs of interest. Another problem is that because the data come from different sources, it must be linked together. This requires an identifying variable (e.g., such as the supervisor's and subordinate's names) that could compromise the anonymity of the respondents and reduce their willingness to participate or change the nature of their responses. In addition, it can also result in the loss of information when data on both the predictor and criterion variables are not obtained. Another disadvantage is that the use of this remedy may require considerably more time, effort, and/or cost on the part of the researcher.

Temporal, proximal, psychological, or methodological separation of measurement. When it is not possible to obtain data from different sources, another potential remedy is to separate the measurement of the predictor and criterion variables. This might be particularly important in the study of attitude–attitude relationships. This separation of measurement can be accomplished in several ways. One is to create a temporal separation by introducing a time lag between the measurement of the predictor and criterion variables. Another is to create a psychological separation by using a cover story to make it appear that the measurement of the predictor variable is not connected with or related to the measurement of the criterion variable. Still another technique is to proximally or methodologically separate the measures by having re-

spondents complete the measurement of the predictor variable under conditions or circumstances that are different from the ones under which they complete the measurement of the criterion variable. For example, researchers can use different response formats (semantic differential, Likert scales, faces scales, open-ended questions), media (computer based vs. paper and pencil vs. face-to-face interviews), and/or locations (e.g., different rooms or sites) for the measurement of the predictor and criterion variables.

With respect to the response processes discussed earlier, the introduction of a temporal, proximal, or psychological separation between the measurement of the predictor and criterion variables has several beneficial effects. First, it should reduce biases in the retrieval stage of the response process by eliminating the saliency of any contextually provided retrieval cues. Second, it should reduce the respondent's ability and/or motivation to use previous answers to fill in gaps in what is recalled and/or to infer missing details. The temporal separation does this by allowing previously recalled information to leave short-term memory, whereas the locational separation does this by eliminating common retrieval cues and the psychological separation does this by reducing the perceived relevance of the previously recalled information in short-term memory. Third, creating a temporal, proximal, or psychological separation should reduce biases in the response reporting or editing stage of the response process by making prior responses less salient, available, or relevant. This diminishes the respondent's ability and motivation to use his or her prior responses to answer subsequent questions, thus reducing consistency motifs and demand characteristics.

There are, of course, some disadvantages to separating the measurement of the predictor and criterion variables. One is that the separation of the measurement of these variables potentially allows contaminating factors to intervene between the measurement of the predictor and criterion variables. For example, although time lags may help reduce common method biases because they reduce the salience of the predictor variable or its accessibility in memory, if the lag is inordinately long for the theoretical relationship under examination, then it could mask a relationship that really exists. Therefore, the length of the time lag must be carefully calibrated to correspond to the process under examination. In addition, if the time lag is long, then respondent attrition may also become a problem. Similarly, a disadvantage of using a psychological separation to reduce common method biases is that they can permit the intrusion of potentially contaminating factors. Finally, a joint disadvantage of all of these methods of introducing a separation between the measurement of the predictor and criterion variables is that they generally take more time, effort, and expense to implement. Thus, although there are some distinct advantages to introducing a separation in measurement, the use of this technique is not without costs.

Protecting respondent anonymity and reducing evaluation apprehension. There are several additional procedures that can be used to reduce method biases, especially at the response editing or reporting stage. One is to allow the respondents' answers to be anonymous. Another is to assure respondents that there are no right or wrong answers and that they should answer questions as honestly as possible. These procedures should reduce people's evaluation apprehension and make them less likely to edit their responses to be more socially desirable, lenient, acquiescent, and consistent with how they think the researcher wants them to respond. Obviously, the primary disadvantage of response anonymity is that it cannot easily be used in conjunction with the two previously described procedural remedies. That is, if the researcher separates the source or the measurement context of the predictor and criterion variables, he or she must have some method of linking the data together. This compromises anonymity, unless a linking variable that is not related to the respondent's identity is used.

Counterbalancing question order. Another remedy that researchers might use to control for priming effects, item-context-induced mood states, and other biases related to the question context or item embeddedness is to counterbalance the order of the measurement of the predictor and criterion variables.

In principle, this could have the effect of neutralizing some of the method biases that affect the retrieval stage by controlling the retrieval cues prompted by the question context. However, the primary disadvantage of counterbalancing is that it may disrupt the logical flow and make it impossible to use the funneling procedure (progressing logically from general to specific questions) often recommended in the survey research literature (Peterson, 2000).

Improving scale items. Moving beyond issues of the source and context of measurement, it is also possible to reduce method biases through the careful construction of the items themselves. For example, Tourangeau et al. (2000) noted that one of the most common problems in the comprehension stage of the response process is item ambiguity and cautioned researchers to (a) define ambiguous or unfamiliar terms; (b) avoid vague concepts and provide examples when such concepts must be used; (c) keep questions simple, specific, and concise; (d) avoid double-barreled questions; (e) decompose questions relating to more than one possibility into simpler, more focused questions; and (f) avoid complicated syntax. Another way to improve scale items is to eliminate item social desirability and demand characteristics. This can be done by using ratings of the social desirability or demand characteristics of each question to identify items that need to be eliminated or reworded (cf. Nederhof, 1985). Still another way to diminish method biases is to use different scale endpoints and formats for the predictor and criterion measures. This reduces method biases caused by commonalities in scale endpoints and anchoring effects. Finally, research (cf. Tourangeau et al., 2000) suggests that acquiescence bias can be reduced by avoiding the use of bipolar numerical scale values (e.g., –3 to +3) and providing verbal labels for the midpoints of scales.

Although we can think of no disadvantages of reducing item ambiguity, social desirability, and demand characteristics, it may not always be desirable to vary the scale anchors and formats and to avoid the use of bipolar scale values. For example, altering scale anchors can change the meaning of a construct and potentially compromise its validity, and the use of unipolar scale values for constructs that are naturally bipolar in nature may not be conceptually appropriate. Therefore, we would caution researchers to be careful not to sacrifice scale validity for the sake of reducing common method biases when altering the scale formats, anchors, and scale values.

Statistical Remedies

It is possible that researchers using procedural remedies can minimize, if not totally eliminate, the potential effects of common

method variance on the findings of their research. However, in other cases, they may have difficulty finding a procedural remedy that meets all of their needs. In these situations, they may find it useful to use one of the statistical remedies that are available. The statistical remedies that have been used in the research literature to control for common method biases are summarized in Table 4 and are discussed in the section that follows.

Harman's single-factor test. One of the most widely used techniques that has been used by researchers to address the issue of common method variance is what has come to be called Harman's one-factor (or single-factor) test. Traditionally, researchers using this technique load all of the variables in their study into an exploratory factor analysis (cf. Andersson & Bateman, 1997; Aulakh & Gencturk, 2000; Greene & Organ, 1973; Organ & Greene, 1981; Schriesheim, 1979) and examine the unrotated factor solution 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 among the measures. More recently, some researchers using this technique (cf. Iverson & Maguire, 2000; Korsgaard & Roberson, 1995; Mossholder, Bennett, Kemery, & Wesolowski, 1998) have used confirmatory factor analysis (CFA) as a more sophisticated test of the hypothesis that a single factor can account for all of the variance in their data.

Despite its apparent appeal, there are several limitations of this procedure. First, and most importantly, although the use of a single-factor test may provide an indication of whether a single factor accounts for all of the covariances among the items, this procedure actually does nothing to statistically control for (or partial out) method effects. If anything, it is a diagnostic technique for assessing the extent to which common method variance may be a problem. However, even on this count, it is an insensitive test. If only one factor emerges from the factor analysis and this factor accounts for all of the variance in the items, it might be reasonable to conclude that common method variance is a major problem (although one could also conclude that the measures of the constructs lacked discriminant validity, were correlated because of a causal relationship, or both). However, in our experience, it is unlikely that a one-factor model will fit the data. It is much more likely that multiple factors will emerge from the factor analysis, and, contrary to what some have said, this is not evidence that the measures are free of common method variance. Indeed, if it were, then it would mean that common method variance would have to completely account for the covariances among the items for it to be regarded as a problem in a particular study. Clearly, this assumption is unwarranted. Therefore, despite the fact this procedure is widely used, we do not believe it is a useful remedy to deal with the problem and turn our attention to other statistical remedies that we feel are better suited for this purpose.

Partial correlation procedures designed to control for method biases. One statistical procedure that has been used to try to control the effects of method variance is the partial correlation procedure. As indicated in Table 4, there are several different variations of this procedure, including (a) partialling out social desirability or general affectivity, (b) partialling out a "marker" variable, and (c) partialling out a general factor score. All of these techniques are similar in that they use a measure of the assumed source of the method variance as a covariate in the statistical analysis. However, they differ in the terms of the specific nature of the source and the extent to which the source can be directly measured. The advantages and disadvantages of each of these techniques is discussed in the paragraphs that follow.

As noted earlier, two variables frequently assumed to cause common method variance are the respondents' affective states and the tendency to respond in a socially desirable manner. In view of this, some researchers (cf. Brief et al., 1988; Burke et al., 1993; Chen & Spector, 1991; Jex & Spector, 1996) have attempted to control for these biases by measuring these variables directly and then partialling their effects out of the predictor and criterion variables. The advantage of this procedure is that it is relatively easy and straightforward to use, in that it only requires that the researcher obtain a measure of the presumed cause of the method biases (e.g., social desirability, negative affectivity) and compare the differences in the partial correlation between the predictor and criterion variables with their zero-order correlation using Olkin and Finn's (1995) significance test (cf. Spector, Chen, & O'Connell, 2000).

However, despite the advantages, there are some limitations of this procedure as well. As noted by Williams, Gavin, and Williams (1996, p. 89), the first limitation of this technique is that the procedure

> does not distinguish between the measures of a construct and the construct itself. As such, the analysis does not incorporate a model of the measurement process. Consequently, it is not possible to assess whether [the directly measured variable] is acting as a measurement contaminant or whether it has a substantive relationship with the [constructs] of interest . . . . Thus, the shared variance attributable to [the directly measured variable] in some past research . . . may reflect some combination of measurement and substantive issues.

Generally speaking, the techniques used to control for common method variance should reflect the fact that it is expected to have its effects at the item level rather than at the construct level. However, for certain types of biases (e.g., social desirability, negative affectivity), it may make theoretical sense to also model the effects of method variance at the construct level (cf. Brief et al., 1988; Williams et al., 1996). Thus, a limitation of this technique is that it prevents a researcher from examining the relative impact of these two distinct types of effects.

Williams et al. (1996) have noted that another limitation of this procedure is that it assumes that the variance shared between the predictor variable of interest, the dependent variable of interest, and the common method variable included in the study is not also shared with some other variables. For example, if the variance shared among the independent variable of interest, the criterion variable of interest, and the common method variable is also shared with other variables included in the study, then the differences between the zero-order relationships and the partialled relationships that are attributed to the common method variable may actually be the result of the other variables that are correlated with them. However, it is important to recognize that the possible impact of "third variables" is a limitation that applies to every method of controlling common method variance and indeed to virtually every modeling technique.

Finally, it is also important to note that this procedure only controls for that portion of common method variance that is

Table 4
Types of Statistical Remedies Used to Address Concerns Regarding Common Method Biases

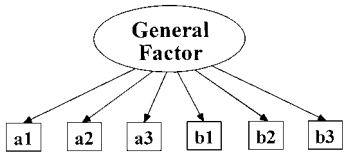
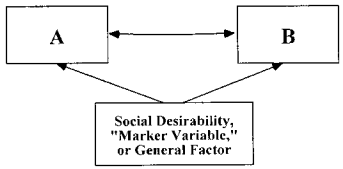
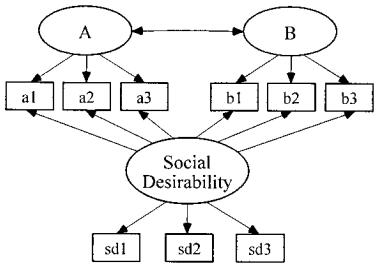
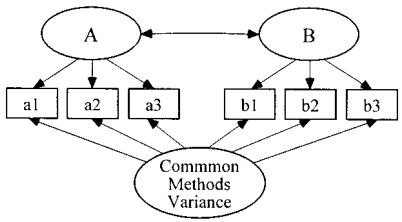
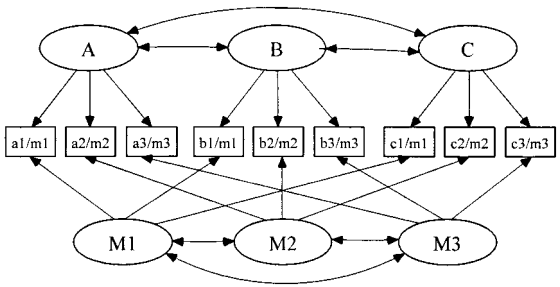
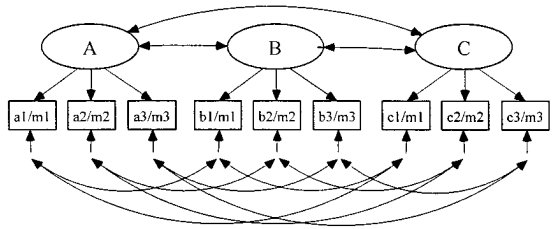
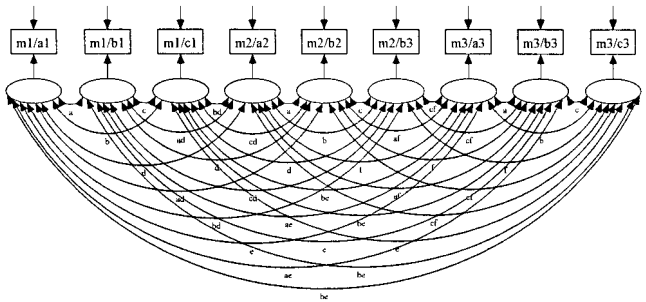
Technique	Description of technique	Example of model	Potential problems
Harman's single-factor test	Include all items from all of the constructs in the study into a factor analysis to determine whether the majority of the variance can be accounted for by one general factor.		Procedure does not statistically control for common method variance. There are no specific guidelines on how much variance the first factor should extract before it is considered a general factor. The likelihood of obtaining more than one factor increases as the number of variables examined increases, thus making the procedure less conservative as the number of variables increases.
Partial correlation procedure	Partialling out social desirability or affectivity as a surrogate for method variance Measures of social desirability (or positive or negative affectivity) are used as surrogates of common methods variance, and the structural parameters are examined both with and without these measures to determine their potential effects on the observed relationships.		Partialling out social desirability or affectivity as a surrogate for method variance Assumes that all of the common methods variance is attributable to the specific surrogate being measured (e.g., social desirability, negative affectivity, positive affectivity). However, these surrogates may not adequately tap the whole common methods variance domain. There is no statistical test to determine differential fit of the model with and without social desirability. Does not permit the researcher to determine whether social desirability serves as a confounding variable at the measurement level or has a substantive effect on the relationships examined. Ignores measurement error.
	Partialling out an unrelated "marker variable" as a surrogate for method variance A "marker variable" that is theoretically unrelated to the constructs of interest is included in the study, and the structural parameters are examined both with and without this measure to determine its potential effects on the observed relationships.		Partialling out an unrelated "marker variable" as a surrogate for method variance Fails to control for some of the most powerful causes of common method biases (e.g., implicit theories, consistency motif, social desirability). Assumes that common method biases have the same effect on all observed variables. Assumes common method variance can only inflate not deflate, the observed relationship between predictor and criterion variable. Ignores measurement error.
	Partialling out a general methods factor After the first unrotated (general) factor is identified using factor analysis, its effects are partialled out to determine whether the structural relationships between the variables of interest are still significant.		Partialling out a general methods factor Although the assumption is that the general factor that is partialled out is composed of common methods, we may also be partialling out some actual covariance between the constructs of interest. Partialling out methods variance in this fashion may produce biased parameter estimates. There is no statistical test to determine differential fit of the model with and without the general factor. Conclusions based on this procedure are sensitive to the number of variables included in the analysis. Ignores measurement error.

Table 4 (*continued*)

Technique	Description of technique	Example of model	Potential problems
Controlling for the effects of a directly measured latent methods factor	Items are allowed to load on their theoretical constructs, as well as on a latent method factor that has its own measurement component, and the significance of the structural parameters is examined both with and without the latent methods factor in the model. The latent methods factor in this case is typically assessed by a surrogate measure (e.g., social desirability, negative or positive affectivity) that is assumed to represent common methods variance.		Assumes that researcher can identify all of the potential sources of common methods bias and that valid measures of these method biases exist. Assumes that method factor does not interact with the predictor and criterion constructs.
Controlling for the effects of an unmeasured latent methods factor	Items are allowed to load on their theoretical constructs, as well as on a latent common methods variance factor, and the significance of the structural parameters is examined both with and without the latent common methods variance factor in the model. In this way, the variance of the responses to a specific measure is partitioned into three components: (a) trait, (b) method, and (c) random error.		Does not allow the researcher to identify the specific cause of the method variance. Potential problems may be encountered with identification of the model. Assumes that method factor does not interact with the predictor and criterion constructs.
Multiple method factors	CFA of MTMM model The most common example of this type of model is the MTMM model, where measures of multiple traits using multiple methods are obtained. In this way, the variance of the responses to a specific measure is partitioned into three components: (a) trait, (b) method, and (c) random error, which permits the researcher to control for both method variance and random error when examining relationships between the predictor and criterion variables.		CFA of MTMM model Potential problems may be encountered with identification of the model. Assumes that method factor does not interact with the predictor and criterion constructs.

(*table continues*)

Table 4 (*continued*)

Technique	Description of technique	Example of model	Potential problems
	Correlated uniqueness model In this model, each observed variable is caused by only one trait factor and a measurement error term. There are no method factors. The model accounts for method effects by allowing the error terms of constructs measured by the same method to be correlated.		Correlated uniqueness model The standard errors of the parameter estimates in this model are biased downward and should not be used for statistical tests. Model assumes that the various method biases are not correlated with each other. Assumes that method factor does not interact with the predictor and criterion constructs.
	Direct product model Unlike the confirmatory factor analysis and correlated uniqueness models, this model assumes that the trait measures interact multiplicatively with the methods of measurement to influence each observed variable. More specifically, this model assumes that the stronger the correlation between traits, the more the intercorrelation between the traits will be influenced by shared method biases.	 Note: Labels associated with each path represent nonlinear constraints that must be placed on the estimates (see Bechger, 1998).	Direct product model The conceptual nature of Trait × Method interactions has not been well articulated, thus making it difficult to predict when they are likely to occur. The effects of the trait and method components cannot be separated in the analysis. Cannot test for relationships between latent trait constructs while simultaneously controlling for method and Trait × Method effects. Does not test for interactions after first controlling for main effects.

Note. CFA = confirmatory factor analysis; MTMM = multitrait–multimethod.

attributable to the specific surrogate being measured (e.g., social desirability, positive or negative affectivity). However, given the wide variety of possible causes of method variance discussed in our article, this technique cannot be regarded as a complete solution to the problem.

Another partial correlation technique that has been recently recommended is the use of a marker variable to control for common method biases (Lindell & Brandt, 2000; Lindell & Whitney, 2001). Lindell and his colleagues have argued that if a variable can be identified on theoretical grounds that should not be related to at least one other variable included in the study, then it can be used as a marker in that any observed relationships between it and any of the other variables can be assumed to be due to common method variance. Moreover, they conclude that partialling out the average correlation between the marker variable and the other variables included in the study should allow the researcher to control for the possible contaminating effect of method biases.

The principal advantage of this procedure is its ease of implementation, especially if one uses the smallest observed correlation among the manifest variables as a proxy for common method variance, like Lindell and Brandt (2000). However, the marker variable technique to control for common method variance has a number of conceptual and empirical problems. From a conceptual point of view, a major problem is that this procedure fails to control for some of the most powerful causes of common method biases (e.g., implicit theories, consistency motif, social desirability). Because a marker variable is one that most people believe should not be related to the predictor or criterion variable, there is no reason to expect that it provides an estimate of the effect of a person's implicit theory about why the predictor and criterion variables should be related, and therefore partialling out the effects of the marker variable will not control for these sources of common method variance. For example, few people would argue on theoretical grounds that an employee's self-reported shoe size should be related to either an employee's performance or the employee's ratings of his or her supervisor's supportive leadership behavior. Therefore, self-reported shoe size meets all of the criteria identified by Lindell and Whitney (2001) and would be an excellent marker variable. However, it is hard to imagine how the strength of the relationship between reported shoe size and employee performance could possibly represent an employees' implicit theory about why supportive leader behavior should be related to employee performance, and therefore partialling out the effects of shoe size would not control for this source of common method variance. The same could be said for the ability of shoe size to serve as a marker variable to control for consistency motif and/or social desirability.

A final conceptual problem with this technique is that it assumes that the common method factor represented by the marker variable "has exactly the same impact on all of the observed variables" (Lindell & Whitney, 2001, p. 116). Although we tend to agree with Lindell and Whitney (2001) that this might be a reasonable assumption for some sources of common method variance (e.g., common scale format, common anchors, or leniency biases), there is no reason why this should necessarily be true for other types of method biases. For example, there is no reason to assume that implicit theories would affect the relationships between all pairs of measures in the same way. In any case, because some of the other methods that control for common method variance do not have to make this tenuous assumption, this should be regarded as a relative disadvantage of this approach.

In addition to these conceptual problems, there are also three important empirical problems with the use of the marker variable technique. First, the procedure is based on the assumption that common method variance can only inflate, not deflate, the observed relationship between a predictor and criterion variable (see Lindell & Whitney, 2001, p. 115). However, as noted earlier, common method variance can inflate, deflate, or have no effect on the observed relationships between predictors and criterion variables (cf. Cote & Buckley, 1988). Second, the procedure ignores measurement error. Third, it assumes that common method factors do not interact with traits—a point that has been disputed by Campbell and O'Connell (1967, 1982), Bagozzi and Yi (1990), and Wothke and Browne (1990), among others.

The last of the partial correlation procedures used in previous research is the general factor covariate technique (Bemmels, 1994; Dooley & Fryxell, 1999; Organ & Greene, 1981; Parkhe, 1993; Podsakoff & Todor, 1985). In this procedure, the first step is to conduct an exploratory factor analysis of the variables included in the study. Then, a scale score for the first unrotated factor (which is assumed to contain the best approximation of common method variance) is calculated and partialled out of the relationship between the predictor and criterion variable. This procedure shares some of the same advantages and disadvantages as the other two partial correlation procedures. The principal advantage is that it is relatively easy to use because the researcher does not have to identify the specific source of the common method variance. However, like some of the partial correlation procedures, it ignores measurement error. In addition, another important disadvantage is that this general method factor may reflect not only common method variance among the measures of the constructs but also variance due to true causal relationships between the constructs. Indeed, it is impossible to separate these two sources of variation using this technique. Consequently, covarying out the effects of this general factor score may produce biased parameter estimates of the relationship between the constructs of interest (cf. Kemery & Dunlap, 1986).

Controlling for the effects of a directly measured latent methods factor. Up to this point, none of the statistical methods discussed are able to adequately account for measurement error or distinguish between the effects of a method factor on the measures of the construct and the construct itself. To address these issues, researchers have turned to the use of latent variable models. One approach that has been used involves directly measuring the presumed cause of the method bias (e.g., social desirability, negative affectivity, or positive affectivity), modeling it as a latent construct, and allowing the indicators of the constructs of interest to load on this factor as well as their hypothesized constructs (see figure with social desirability as the latent methods factor in Table 4). For example, Williams has used this approach to examine the potentially biasing effects of negative affectivity on the relationships between a variety of job attitudes and role perceptions (Williams et al., 1996) and the effects of positive and negative emotionality (Williams & Anderson, 1994). The advantages of using this approach are that it (a) allows measurement error in the method factor to be estimated, (b) models the effects of the biasing factor on the measures themselves rather than directly on the theoretical constructs of interest, and (c) does not constrain the effects of the methods factor on the individual measures to be equal.

However, there are disadvantages to this approach as well. To use this method, the researcher must know what the most important sources of method biases are in his or her study and be able to directly measure them. This is a serious limitation because it is often difficult to identify the most important sources of method bias in a given situation. In addition, valid measures for some of the sources of bias that a researcher might identify simply do not exist (e.g., implicit theories, consistency biases, common scale format and anchors). Finally, this technique requires the assumption that the method factor does not interact with the constructs of interest. As previously noted, this assumption has been questioned by Campbell and O'Connell (1967), Bagozzi and Yi (1990), and Wothke and Browne (1990).

Controlling for the effects of a single unmeasured latent method factor. Another latent variable approach that has been used involves adding a first-order factor with all of the measures as indicators to the researcher's theoretical model. Such a model is depicted in Table 4 and has been used in a number of studies (e.g., Carlson & Kacmar, 2000; Carlson & Perrewe, 1999; Conger, Kanungo, & Menon, 2000; Elangovan & Xie, 1999; Facteau, Dobbins, Russell, Ladd, & Kudisch, 1995; MacKenzie, Podsakoff, & Fetter, 1991, 1993; MacKenzie, Podsakoff, & Paine, 1999; Moorman & Blakely, 1995; Podsakoff & MacKenzie, 1994; Podsakoff, MacKenzie, Moorman, & Fetter, 1990). One of the main advantages of this technique is that it does not require the researcher to identify and measure the specific factor responsible for the method effects. In addition, this technique models the effect of the method factor on the measures rather than on the latent constructs they represent and does not require the effects of the method factor on each measure to be equal.

Like the other methods discussed, there are also some disadvantages of this approach. One is that although this technique controls for any systematic variance among the items that is independent of the covariance due to the constructs of interest, it does not permit the researcher to identify the specific cause of the method bias. Indeed, the factor may reflect not only different types of common method variance but also variance due to relationships between the constructs other than the one hypothesized. Another disadvantage is that if the number of indicators of the constructs is small relative to the number of constructs of interest, the addition of the method factor can cause the model to be underidentified. As a solution to this problem, some researchers have constrained the measurement factor loadings to be equal. Although this solves the identification problem, it also undermines one of the advantages of this technique. A final disadvantage is that this technique assumes that the method factor does not interact with trait factors. Thus, this approach does have some limitations even though it is attractive because it does not require the researcher to identify the potential source of the method variance in advance.

Use of multiple-method factors to control method variance. A variation of the common method factor technique has been frequently used in the literature (cf. Bagozzi & Yi, 1990; Cote & Buckley, 1987; Williams et al., 1989). This model differs from the common method factor model previously described in two ways. First, multiple first-order method factors are added to the model. Second, each of these method factors is hypothesized to influence only a subset of the measures rather than all of them. The most common example of this type of model is the multitrait–multimethod (MTMM) model (Campbell & Fiske, 1959), where measures of multiple traits using multiple methods are obtained. In this way, the variance of the responses to a specific measure can be partitioned into trait, method, and random error components, thus permitting the researcher to control for both method variance and random error when looking at relationships between the predictor and criterion variables. The first advantage of this technique is that it allows the researcher to examine the effects of several methods factors at one time. A second advantage is that this technique permits the researcher to examine the effects of specifically hypothesized method biases by constraining some of the paths from the method factors to measures that they are not hypothesized to influence to be equal to zero. A third advantage of this technique is that it does not require the direct measurement of these hypothesized method biases.

Despite these advantages, the principal disadvantages of this technique are (a) potentially severe problems may be encountered when estimating these models because of identification problems, specification errors, and sampling errors (Spector & Brannick, 1995); (b) it assumes that the method factors do not interact with the predictor and criterion constructs; and (c) it requires the researcher to identify the potential sources of method bias to specify the relationships between the method factors and the measures.

Correlated uniqueness model. One technique that has been recommended to address the estimation problems encountered with the use of the MTMM model is the correlated uniqueness model (Kenny, 1979; Marsh, 1989; Marsh & Bailey, 1991). In an MTMM model, each observed variable is modeled as being caused by one trait factor, one method factor, and one measurement error term. However, in a correlated uniqueness model, each observed variable is caused by only a trait factor and a measurement error term. There are no method factors. Instead, the correlated uniqueness model accounts for method effects by allowing the error terms of variables measured by the same method to be correlated. Thus, like the MTMM model, this technique requires the researcher to be able to identify the sources of method variance so that the appropriate pattern of measurement error correlations can be estimated.

The principal advantage of the correlated uniqueness approach is that it is more likely than the MTMM model to converge and to produce proper parameter estimates (cf. Becker & Cote, 1994; Conway, 1998; Marsh, 1989; Marsh & Bailey, 1991). In addition, like the CFA technique, the correlated uniqueness technique (a) allows the researcher to examine the effects of multiple method biases at one time, (b) permits the researcher to examine the effects of specifically hypothesized method biases, and (c) does not require the direct measurement of these hypothesized method biases. Historically, its principal disadvantage was its inability to estimate the proportion of variance in a measure caused by method effects, although recent developments by Conway (1998) and Scullen (1999) have demonstrated how this can be done by averaging the measurement error correlations. However, other serious disadvantages remain. As noted by Lance, Noble, and Scullen (2002), these include the fact that (a) method effects are constrained to be orthogonal under the correlated uniqueness model; (b) the estimates of the trait variance components are likely to be biased because the error terms in this model are an aggregation of systematic, nonsystematic, and method effects; (c) this model assumes that the various method biases are uncorrelated; and (d) this method assumes that trait and method effects do not interact.

Direct product model. A common criticism of all of the statistical remedies involving latent constructs that have been discussed so far is that they require the assumption that the method factors do not interact with the predictor and criterion constructs (i.e., trait factors). According to Campbell and O'Connell (1967, 1982), this is a tenuous assumption because there are circumstances in which trait and method factors are likely to interact multiplicatively such that method biases augment (or attenuate) the correlation between strongly related constructs more than they augment (or attenuate) the correlations between weakly related constructs. The direct product model (Bagozzi & Yi, 1990; Bechger, 1998; Browne, 1984; Wothke & Browne, 1990) was developed to take Trait × Method interactions like these into account, and it does this through a radically different underlying model structure that replaces the additive paths in a traditional MTMM CFA with relationships that reflect multiplicative interactions between traits and methods. To the extent that Trait × Method interactions are present, this is obviously a major advantage. In addition, like the correlated uniqueness model, the direct product model is more likely than the MTMM model to converge and produce proper solutions.

However, these advantages are undermined by some fairly serious disadvantages. One major disadvantage is that the direct product model cannot provide separate estimates of the amount of trait and method variance present in a measure, thus making it difficult to assess item validity by examining how well each measure reflects the trait it is intended to represent. A second disadvantage is that the direct product model does not control for the main effects of the trait and method factors when testing the Trait × Method interactions, as is standard procedure when testing interactions. Third, although this technique tests for method biases, it cannot be used to statistically control for them while simultaneously estimating the relationship between a predictor and criterion construct. Still other disadvantages include that it (a) does not permit researchers to identify the specific cause of any method biases observed and (b) requires the specification of a complex set of equality constraints when estimated using the most widely available structural equation modeling programs (e.g., LISREL, EQS, AMOS). However, in our view, perhaps the most serious disadvantage is that Trait × Method interactions may not be very common or potent. Indeed, the weight of the evidence (cf. Bagozzi & Yi, 1990, 1991; Bagozzi et al., 1991; Becker & Cote, 1994; Hernandez & Gonzalez-Roma, 2002; Kumar & Dillon, 1992) suggests that although Trait × Method interactions are theoretically possible and do exist in certain circumstances, they are not usually very strong and a simpler MTMM model may be just as good as the more complicated direct product approach.

Comparison of Statistical Remedies for Common Method Biases

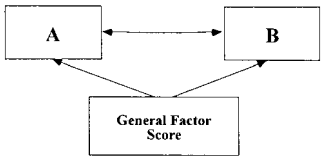
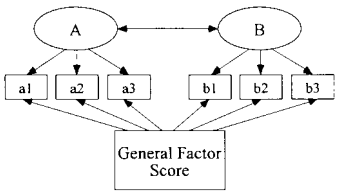
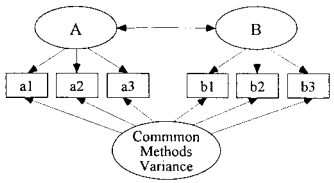
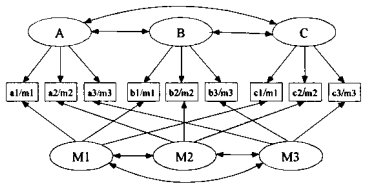
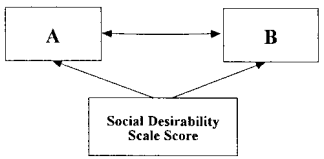
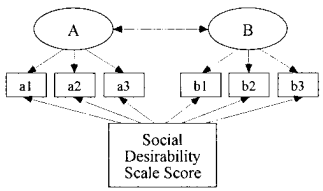
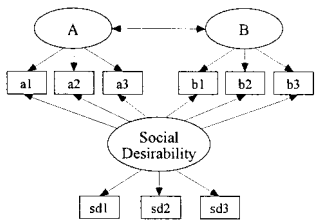
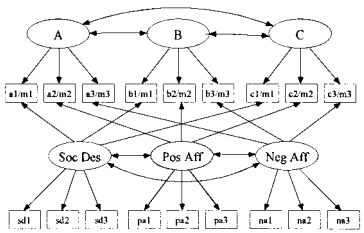
The preceding section provides a review of the statistical procedures that have been used in the literature to control for common method biases. However, there are some potential remedies that have not been tried, which, when combined with the methods that have been used, suggest a continuum of ways to statistically control for common method biases. Table 5 summarizes this set of potential remedies and highlights some of their key features. As shown in Table 5, some of the statistical approaches require the researcher to identify the source of method bias and require a valid measure of the biasing factor whereas others do not. In addition, the approaches differ in the extent to which they control a variety of potential problems, including (a) the ability to distinguish method bias at the measurement level from method bias at the construct level, (b) measurement error, (c) single versus multiple sources of method bias, and (d) Trait × Method interactions. Generally speaking, as one moves from left to right and from top to bottom in the table, the approaches require more effort on the part of the researcher, either because the model specification is more complex or because additional measures of the method bias must be obtained.

The approaches shown in the second column of the table (partial correlation approaches) have the advantage that they are relatively easy to implement and can be used even with small sample sizes (i.e., samples too small to meet the requirements for latent variable structural equation modeling). However, they are the weakest among the statistical remedies because they (a) fail to distinguish method bias at the measurement level from method bias at the construct level, (b) ignore measurement error in the method factor, (c) only control for a single source of method bias at a time, and (d) ignore Method × Trait interactions. Thus, although fairly easy to implement and widely used, they are not very satisfactory methods for controlling for method biases.

The second set of approaches (single-method-scale-score approaches) are an improvement over the first group of techniques because they distinguish method bias at the measurement level from method bias at the construct level and appropriately model it at the measurement level. To our knowledge, no one has used this set of procedures to control for method variance, probably because these techniques still ignore measurement error in the method factor, even though there is no reason to do so. One situation in which the technique identified in Cell 2B might be useful is where the measure of the potential biasing factor (e.g., social desirability) has a large number of items (e.g., 10 or more) and/or a complex factor structure. In this instance, the researcher may decide to use a scale score to represent the source of bias because if the measures are all treated as separate indictors of the biasing factor, then any nonhypothesized measurement error covariances will contribute to the lack of fit of the model. If these measurement error covariances are substantial, then the goodness-of-fit of the model will be disproportionately influenced by them. However, even in this case, we believe that a better remedy to this problem would be to create a small number of testlets (Bagozzi & Heatherton, 1994) composed of random subsets of the measures of the biasing factor and then use the procedure in Cell 3B.

The third set of approaches (single-method-factor approaches) have the advantages of estimating method biases at the measurement level and controlling measurement error. Perhaps because of these advantages, these techniques have been frequently used in the literature. Examples of the use of the single-common-method-factor approach shown in Cell 3A include Carlson and Kacmar (2000), Elangovan and Xie (1999), MacKenzie et al. (1991, 1993), and Podsakoff et al. (1990) and examples of the use of the single-specific-method-factor approach shown in Cell 3B include Williams and Anderson (1994) and Williams et al. (1996). The main disadvantages of these approaches are that they only control for a single source of method bias at a time and assume that

Table 5
Summary of Statistical Remedies Used to Control Common Method Biases

Requirements	Partial correlation approaches	Single-method-scale-score approaches	Single-method-factor approaches	Multiple-method-factor approaches
	Disadvantages Fails to distinguish method bias at the measurement level from method bias at the construct level. Ignores measurement error in the method factor. Only controls for a single source of method bias at a time. Ignores Method × Trait interactions.	Disadvantages Ignores measurement error in the method factor. Only controls for a single source of method bias at a time. Ignores Method × Trait interactions.	Disadvantages Only controls for a single source of method bias at a time. Ignores Method × Trait interactions.	Disadvantage Ignores Method × Trait interactions.
Does not require the researcher to identify the precise source of method bias. Does not require a valid measure of the biasing factor.	1A 	2A 	3A 	4A 
Requires the researcher to identify the precise source of method bias. Requires a valid measure of the biasing factor.	1B 	2B 	3B 	4B 

Method × Trait interactions are not present. How serious these disadvantages are depends on how confident the researcher is that the method factor adequately captures the main source of method bias and that Method × Trait interactions do not exist. The former is a judgment that has to be made primarily on conceptual grounds. However, on the latter issue, the empirical evidence suggests that Method × Trait interactions are unlikely to be very strong (Becker & Cote, 1994).

The final set of approaches shown in Table 5 (multiple-method-factor approaches) is the strongest of the approaches depicted. Cell 4A represents the classic MTMM model, whereas 4B represents a situation in which the researcher has directly measured several suspected sources of method bias (e.g., social desirability, positive affectivity and negative affectivity), models these biasing factors as latent variables with multiple indicators, and estimates their effects on the measures of the constructs of interest. These approaches are particularly strong because they model method bias at the measurement level, control for measurement error, and incorporate multiple sources of method bias. However, they have two key disadvantages. The first is that because the models are complex, estimation problems may be encountered. It is widely recognized that this is especially a problem for the MTMM model shown in Cell 4A (cf. Becker & Cote, 1994; Brannick & Spector, 1990; Marsh, 1989; Marsh & Bailey, 1991; Spector & Brannick, 1995). It is for this reason that some researchers (Conway, 1998; Kenny, 1979; Marsh, 1989; Marsh & Bailey, 1991; Scullen, 1999) have recommended the use of the correlated uniqueness model. However, because the MTMM model is conceptually and empirically superior to the correlated uniqueness model as long as it is not empirically underidentified (cf. Lance, Noble, & Scullen, 2002), the correlated uniqueness model is not shown in the summary table. The second problem encountered by all of the approaches in the final column of Table 5 is that they do not take Method × Trait interactions into account.

Although not shown in Table 5, one approach that takes Method × Trait interactions into account is the direct product model. Unlike the other approaches, it (a) distinguishes method bias at the measurement level from method bias at the construct level, (b) takes measurement error in the method factor into account, (c) can control for multiple sources of method bias at the same time, and (d) captures Method × Trait interactions. However, at the present time, this approach is not recommended for two reasons. First, the conceptual nature of Trait × Method interactions has never been well articulated, thus making it difficult to predict when these interactions are likely to occur. Second, the existing empirical evidence suggests that these interactions may be fairly rare in research settings, and even when they do occur, they are generally weak and a simpler MTMM model may be just as good in these situations as the more complicated direct product approach (cf. Bagozzi & Yi, 1991; Becker & Cote, 1994). Therefore, additional research is needed on these issues before the use of the direct product model is likely to be widely recommended as a technique for controlling common method biases.

Recommendations for Controlling Method Biases in Research Settings

Our discussion in this article clearly indicates that common method variance can have a substantial impact on the observed relationships between predictor and criterion variables in organizational and behavioral research. Although estimates of the strength of the impact of common method biases vary (cf. Bagozzi & Yi, 1990; Cote & Buckley, 1987; Spector, 1987, 1994; Williams et al., 1989), their average level is quite substantial. Indeed, the evidence reported by Cote and Buckley (1987) from 70 MTMM studies conducted in a variety of disciplines indicates that the observed relationship between a typical predictor and criterion variable is understated by approximately 26% because of common method biases. Moreover, as noted in Table 2, this bias may come from a number of different sources that could be in operation in any given research study. Therefore, we believe that researchers would be wise to do whatever they can to control for method biases, and the method used to control for it should to be tailored to match the specific research setting.

Figure 1 describes a set of procedures that might be used to control for method biases that are designed to match several typical research settings. Generally speaking, we recommend that researchers follow good measurement practice by implementing all of the procedural remedies related to questionnaire and item design (e.g., eliminate item ambiguity, demand characteristics, social desirability). Following this, we recommend that they implement additional procedural and statistical remedies to control for the method biases that are likely to be present in their specific research situation. This can be done by considering four key questions: (a) Can the predictor and criterion variables be obtained from different sources? (b) Can the predictor and criterion variables be measured in different contexts? (c) Can the source of the method bias be identified? and (d) Can the method bias be validly measured?

Beginning at the bottom left of Figure 1 (Situation 1), if the predictor and criterion variables can be measured from different sources, then we recommend that this be done. Additional statistical remedies could be used but in our view are probably unnecessary in these instances.

Moving to Situation 2 in Figure 1, if the predictor and criterion variables are obtained from the same source but can be obtained in a different context, the researcher has a good idea about the source(s) of the method bias (e.g., social desirability, negative affectivity, positive affectivity), and the suspected bias can be validly measured, then we recommend that researchers (a) separate the measures of the predictor and criterion variables (temporally, proximally, psychologically, and/or methodologically) and (b) measure the biasing factor(s) and estimate its effects on the measures using the single-specific-method-factor or multiple-specific-method-factors approaches (Cells 3B or 4B in Table 5). The rationale for this recommendation is that separating the measurement of the predictor and criterion variable should help to prevent method biases because of a common rater, whereas the statistical remedy controls for these biases if they happen to occur in spite of this procedural control.

Situation 3 represents virtually the same circumstances as described in Situation 2, with the exception that there is no valid way to measure the biasing factor suspected to be in operation. In this instance, we recommend that researchers separate the measures of the predictor and criterion variables (temporally, proximally, psychologically, and/or methodologically) and estimate any residual effects of the suspected source of method bias using the single-common-method-factor approach specified in Cell 3A in Table 5

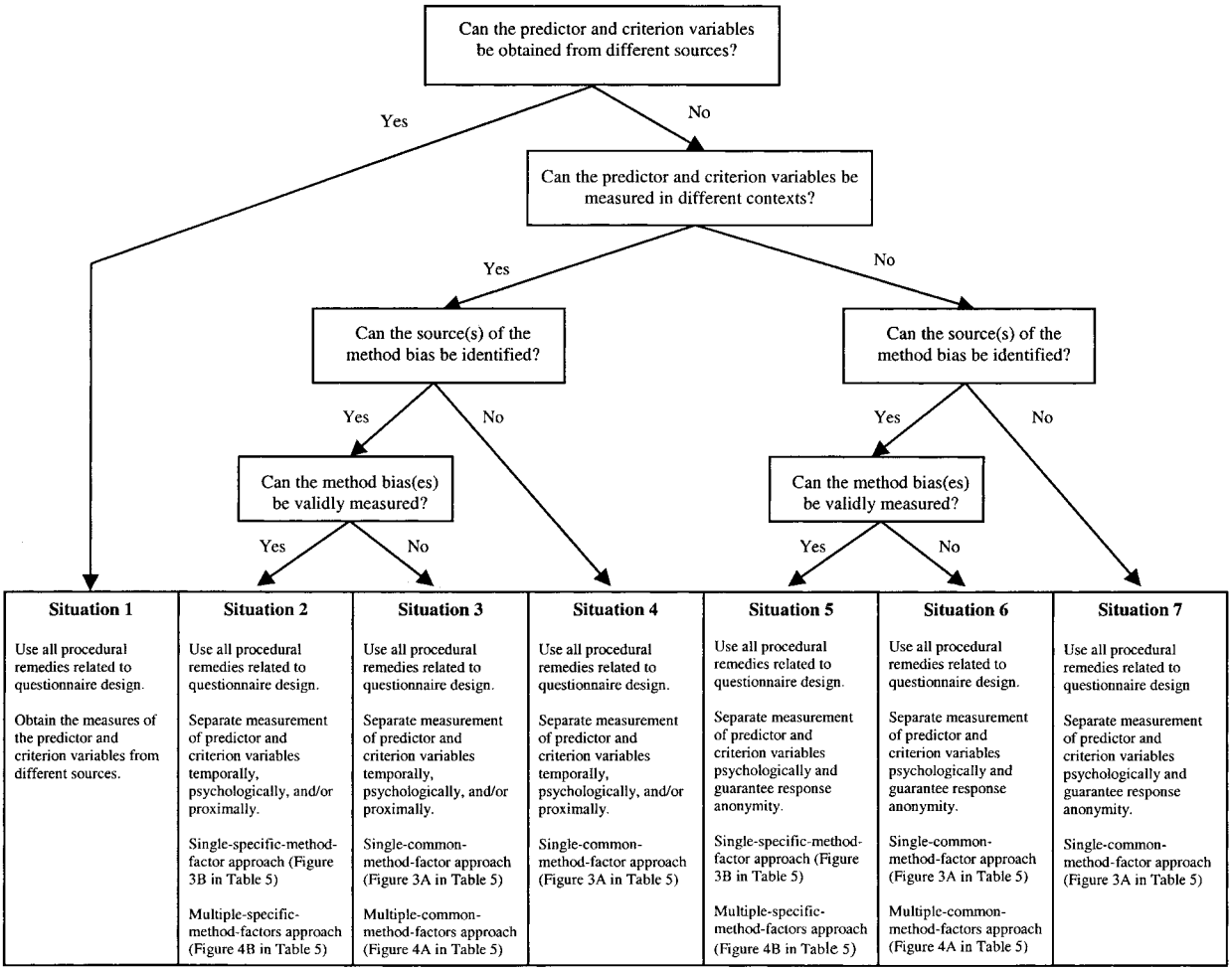


Figure 1. Recommendations for controlling for common method variance in different research settings.

or the multiple-common-method-factors approach specified in Cell 4A (depending on the number of biasing factors believed to be in operation). This situation might arise if leniency biases are expected to be a potential problem in a study because changing the measurement context does not control for this form of bias and there is no valid scale to directly measure leniency biases. Under these circumstances, a method of statistical control that does not require the biasing factor to be directly measured must be used (e.g., single-common-method-factor approach). However, in other situations in which several biasing factors that are not directly measurable exist, it may be necessary to use the approach specified in Cell 4A (multiple-common-method-factors approach).

The recommendations for Situation 4 are the same as Situation 3, with the exception that the multiple-common-method-factors approach (Cell 4A) cannot be used. This method of statistical control cannot be used in this case because the specification of a multiple-common-method-factors model requires the sources of method bias to be known by the researcher; otherwise, it is impossible to specify which method factor influences which specific measures.

Situations 5–7 differ from Situations 2–4 in that it is not possible to measure the predictor and criterion variables in different times or locations. As a result, the procedural remedy of separating the measurement of the predictor and criterion variables temporally or proximally cannot be used. One potential procedural remedy in this case is to attempt to reduce method bias by guaranteeing response anonymity. Another procedural remedy is to separate the predictor and criterion variable psychologically. As noted earlier, this might be done by creating a cover story to make it appear that the measurement of the predictor variable is not connected with or related to the measurement of the criterion variable. However, because guaranteeing anonymity and/or psychological separation does not eliminate all of the different method biases associated with a common rater and measurement context (e.g., dispositional moods states, leniency biases, acquiescence biases, contextually provided retrieval cues), the researcher needs to depend more on the statistical remedies in these cases.

Situation 5 is one in which the predictor and criterion variables cannot be obtained from different sources or contexts, but the source(s) of the method bias can be identified and a valid scale to

measure it exists. Under these circumstances, we recommend that researchers try to statistically control for the effect of these bias(es) using the single-specific-method-factor approach (Cell 3B in Table 5) or the multiple-specific-method-factors approach (Cell 4B in Table 5).

Situation 6 is similar to Situation 5, except that the method bias(es) cannot be validly measured. Under these circumstances, the single-common-method-factor approach (Cell 3A in Table 5) or the multiple-common-method-factors approach (Cell 4A), should be used to statistically control for method biases. For example, if the predictor and criterion measures are obtained from the same source in the same context and the researcher suspects that implicit theories may bias the raters' responses, then the single-common-method-factor approach should be used. Alternatively, in a study that obtained measures of the predictor and criterion variables from three different sources (e.g., peers, supervisors, and subordinates) all at the same time, the multiple-common-method-factors approach should be used.

Finally, Situation 7 displays a circumstance in which a researcher cannot obtain the predictor and criterion variables from different sources, cannot separate the measurement context, and cannot identify the source of the method bias. In this situation, it is best to use a single-common-method-factor approach (Cell 3A in Table 5) to statistically control for method biases.

Of course, it is important to note that it may be impossible to completely eliminate all forms of common method biases in a particular study, and controlling for common method biases becomes more complex in multiple equation systems in which relationships between criterion variables are hypothesized (e.g., mediating effects models). However, the goal should be to reduce the plausibility of method biases as an explanation of the relationships observed between the constructs of interest. In most instances, this involves a combination of the procedural and statistical remedies previously discussed. In general, we recommend that researchers first try to prevent method biases from influencing their results by implementing any of the procedural remedies that make sense within the context of their research. However, because it may be impossible to eliminate some sources of bias with procedural remedies, we recommend that researchers follow this up with appropriate statistical remedies.

For example, a researcher interested in the potential mediating effects of employees' trust in their leaders and commitment to the organization on the relationships between transformational leadership behaviors and employee performance may find it difficult (or even undesirable) to obtain all of the measures from different sources. In this case, even though the researcher might be able to maintain the independence of the leader behavior and performance measures by obtaining the ratings of the leader from the subordinates and the ratings of performance from the leader, there is still the problem of where to obtain the measures of the employee attitudes (trust and commitment) to maintain their independence. Of course, one solution to this dilemma would be to obtain objective measures of employee performance from company records and then have the supervisor provide the ratings of their perceptions of the employees' trust and commitment. However, given that one's behavior does not always correlate strongly with one's attitudes (Fishbein & Ajzen, 1975), it is doubtful whether the supervisors' (or anyone else's) perceptions of employees' attitudes is as good a measure as the employees' own self-reports. Another possible procedural remedy is to have the supervisors provide self-reports of their own transformational leadership behavior and the subordinates provide ratings of their own attitudes. Unfortunately, given the evidence that self-reports of behavior are often considerably different from the reports of others (cf. Harris & Schaubroeck, 1988; Jones & Nisbett, 1972), this remedy has other limitations that are associated with it.

Therefore, in the situation described above, the best course of action would probably be to obtain the measures of employee performance from company records; to obtain the measures of the leader behaviors and employee attitudes from the employees but to separate their measurement temporally, contextually, or psychologically; and to statistically control for same-source biases in the leader behavior and employee attitude measures by adding a single-common-method-factor to the structural equation model used to test the hypothesized relationships. However, the key point to remember is that the procedural and statistical remedies selected should be tailored to fit the specific research question at hand. There is no single best method for handling the problem of common method variance because it depends on what the sources of method variance are in the study and the feasibility of the remedies that are available.

Some Additional Considerations

Our discussion of common method variance up to this point has focused on fairly typical situations in which the measures of a predictor and criterion variable are obtained in a field setting using some form of questionnaire. However, there are two other situations that generally have not been discussed in the method variance literature that also deserve our attention. The first of these is in experimental research settings in which mediating processes are examined and the measures of the potential mediators and/or dependent variables are obtained from the same source. The second area relates to the special problems encountered in trying to statistically control for method biases when using formative indicator measurement models (cf. Bollen & Lennox, 1991). In this section, we discuss each of these issues in turn.

Controlling for Method Variance in Experimental Research Examining Mediated Effects

It is often assumed that experimental studies are immune to the method biases discussed in this article because measures of the independent and dependent variable are not obtained from the same person at the same point in time. However, it is not uncommon in experimental studies for researchers to manipulate an independent variable and obtain measures of a potential mediator as well as the dependent variable. The problem with this is that these two measures are usually obtained from the same subjects at the same point in time. In such cases, method biases contribute to the observed relationship between the mediator and the dependent measure. One possible remedy for this is to control for methods effects using the single-common-method-factor approach discussed in Cell 3A of Table 5. As in the case of field research, experimental studies that can demonstrate that the relationships observed between the variables of interest are significant after controlling for method biases provide more compelling evidence than those that do not.

Controlling for Method Variance in Studies Using Formative Constructs

For most of the constructs measured in behavioral research, the relationship between the measures and constructs is implicitly based on classical test theory that assumes that the variation in the scores on measures of a construct is a function of the true score plus error. Thus, the underlying latent construct is assumed to cause the observed variation in the measures (Bollen, 1989; Nunnally, 1978), and the indicators are said to be reflective of the underlying construct. This assumed direction of causality—from the latent variable to its measures—is conceptually appropriate in many instances but not all. Indeed, it has been recognized now for several decades that, for some constructs, it makes more sense conceptually to view causality flowing from the measures to the construct rather than vice versa (Blalock, 1964; Bollen & Lennox, 1991; MacCallum & Browne, 1993). In these cases, the constructs are often called *composite latent variables* and the measures are said to represent formative indicators of the construct. Researchers (cf. Bollen & Lennox, 1991; Law & Wong, 1999) are now beginning to recognize that many of the most widely used constructs in the field (e.g., job satisfaction, role ambiguity, role conflict, task characteristics) are more accurately represented as formative-indicator constructs than they are as reflective-indicator constructs. Perhaps of more importance, research by Law and Wong (1999) has demonstrated that misspecifying measurement relationships by modeling formative-indicator constructs as if they were reflective-indicator constructs can have serious biasing effects on estimates of the relationships between constructs.

Although not previously recognized in the literature, the distinction between formative-indicator and reflective-indicator measurement models is also important because it complicates the treatment of common method biases. The goal of the statistical control procedures discussed in Table 5 is to obtain an estimate of the relationship between the constructs and measures that partials out the effect of method bias. Unfortunately, if a researcher uses the methods shown in the table (i.e., models the effects of a method factor on the formative measures), he or she will not partial the effect of method bias out of the estimated relationship between the formative measures and the construct. This is true because the method factor does not enter into the equation where the relationship between the formative measures and the construct is estimated. Indeed, from a conceptual point of view, the effects of method bias on formatively measured constructs should be modeled at the construct level rather than the item level. If this is done, the estimated relationships between the formative measures and the formative construct will be independent of method bias. This is consistent with Bollen and Lennox's (1991) conceptual argument that for formative constructs, measurement error resides at the construct rather than the item level. Unfortunately, such a model is not identified because, as noted by MacCallum and Browne (1993), the construct-level error term for a construct with formative measures is only identified when there are paths emanating from the construct to at least two reflectively measured constructs that are independent from each other. This suggests that when formative-indicator constructs are an integral part of a study, researchers must be even more careful than normal in designing their research because procedural controls are likely to be the most effective way to control common measurement biases.

Conclusions

Although the strength of method biases may vary across research contexts, a careful examination of the literature suggests that common method variance is often a problem and researchers need to do whatever they can to control for it. As we have discussed, this requires carefully assessing the research setting to identify the potential sources of bias and implementing both procedural and statistical methods of control. Although we clearly have not resolved all of the issues about this important topic, hopefully we have provided some useful suggestions and, perhaps more importantly, a framework that researchers can use when evaluating the potential biasing effects of method variance in their research.

References

Andersson, L. M., & Bateman, T. S. (1997). Cynicism in the workplace: Some causes and effects. *Journal of Organizational Behavior, 18,* 449–469.

Aulakh, P. S., & Gencturk, E. F. (2000). International principal–agent relationships—control, governance and performance. *Industrial Marketing Management, 29,* 521–538.

Bagozzi, R. P., & Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling, 1,* 35–67.

Bagozzi, R. P., & Yi, Y. (1990). Assessing method variance in multitrait–multimethod matrices: The case of self-reported affect and perceptions at work. *Journal of Applied Psychology, 75,* 547–560.

Bagozzi, R. P., & Yi, Y. (1991). Multitrait–multimethod matrices in consumer research. *Journal of Consumer Research, 17,* 426–439.

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly, 36,* 421–458.

Bechger, T. M. (1998). Browne's composite direct product model for MTMM correlation matrices. *Structural Equation Modeling, 5,* 191–204.

Becker, T. E., & Cote, J. A. (1994). Additive and multiplicative method effects in applied psychological research: An empirical assessment of three models. *Journal of Management, 20,* 625–641.

Bemmels, B. (1994). The determinants of grievance initiation. *Industrial and Labor Relations Review, 47,* 285–301.

Berman, J. S., & Kenny, D. A. (1976). Correlational bias in observer ratings. *Journal of Personality and Social Psychology, 34,* 263–273.

Blalock, H. M., Jr. (1964). *Causal inferences in nonexperimental research.* New York: W. W. Norton.

Blaney, P. H. (1986). Affect and memory: A review. *Psychological Bulletin, 99,* 229–246.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110,* 305–314.

Bouchard, T. J., Jr. (1976). Field research methods: Interviewing, questionnaires, participant observation, systematic observation, unobtrusive measures. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 363–414). New York: Wiley.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36,* 129–148.

Brannick, M. T., & Spector, P. E. (1990). Estimation problems in the block-diagonal model of the multitrait–multimethod matrix. *Applied Psychological Measurement, 14,* 325–339.

Brief, A. P., Burke, M. J., George, J. M., Robinson, B. S., & Webster, J.

(1988). Should negative affectivity remain an unmeasured variable in the study of job stress? *Journal of Applied Psychology, 73,* 191–198.

Browne, M. W. (1984). The decomposition of multitrait–multimethod matrices. *The British Journal of Mathematical and Statistical Psychology, 37,* 1–21.

Burke, M. J., Brief, A. P., & George, J. M. (1993). The role of negative affectivity in understanding relations between self-reports of stressors and strains: A comment on the applied psychology literature. *Journal of Applied Psychology, 78,* 402–412.

Campbell, D. T., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Campbell, D. T., & O'Connell, E. J. (1967). Methods factors in multitrait–multimethod matrices: Multiplicative rather than additive? *Multivariate Behavioral Research, 2,* 409–426.

Campbell, D. T., & O'Connell, E. J. (1982). Methods as diluting trait relationships rather than adding irrelevant systematic variance. In D. Brinberg & L. Kidder (Eds.), *Forms of validity in research* (pp. 93–111). San Francisco: Jossey-Bass.

Cannell, C., Miller, P., & Oksenberg, L. (1981). Research on interviewing techniques. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 389–437). San Francisco: Jossey-Bass.

Carlson, D. S., & Kacmar, K. M. (2000). Work–family conflict in the organization: Do life role values make a difference? *Journal of Management, 26,* 1031–1054.

Carlson, D. S., & Perrewe, P. L. (1999). The role of social support in the stressor–strain relationship: An examination of work–family conflict. *Journal of Management, 25,* 513–540.

Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology, 73,* 193–204.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlations as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 75,* 271–280.

Chen, P. Y., & Spector, P. E. (1991). Negative affectivity as the underlying cause of correlations between stressors and strains. *Journal of Applied Psychology, 76,* 398–407.

Collins, W. (1970). Interviewers' verbal idiosyncrasies as a source of bias. *Public Opinion Quarterly, 34,* 416–422.

Conger, J. A., Kanungo, R. N., & Menon, S. T. (2000). Charismatic leadership and follower effects. *Journal of Organizational Behavior, 21,* 747–767.

Conway, J. M. (1998). Understanding method variance in multitrait–multirater performance appraisal matrices: Examples using general impressions and interpersonal affect as measured method factors. *Human Performance, 11,* 29–55.

Cote, J. A., & Buckley, R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research, 24,* 315–318.

Cote, J. A., & Buckley, R. (1988). Measurement error and theory testing in consumer research: An illustration of the importance of construct validation. *Journal of Consumer Research, 14,* 579–582.

Crampton, S., & Wagner, J. (1994). Percept–percept inflation in microorganizational research: An investigation of prevalence and effect. *Journal of Applied Psychology, 79,* 67–76.

Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement, 6,* 475–494.

Cronbach, L. J. (1950). Further evidence on response sets and test validity. *Educational and Psychological Measurement, 10,* 3–31.

Crowne, D., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence.* New York: Wiley.

Dooley, R. S., & Fryxell, G. E. (1999). Attaining decision quality and commitment from dissent: The moderating effects of loyalty and competence in strategic decision-making teams. *Academy of Management Journal, 42,* 389–402.

Eden, D., & Leviatin, V. (1975). Implicit leadership theory as a determinant of the factor structure underlying supervisory behavior scales. *Journal of Applied Psychology, 60,* 736–740.

Elangovan, A. R., & Xie, J. L. (1999). Effects of perceived power of supervisor on subordinate stress and motivation: The moderating role of subordinate characteristics. *Journal of Organizational Behavior, 20,* 359–373.

Facteau, J. D., Dobbins, G. H., Russell, J. E. A., Ladd, R. T., & Kudisch, J. D. (1995). The influence of general perceptions of training environment on pretraining motivation and perceived training transfer. *Journal of Management, 21,* 1–25.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research.* Reading, MA: Addison-Wesley.

Fiske, D. W. (1982). Convergent–discriminant validation in measurements and research strategies. In D. Brinbirg & L. H. Kidder (Eds.), *Forms of validity in research* (pp. 77–92). San Francisco: Jossey-Bass.

Fowler, F. J. (1992). How unclear terms affect survey data. *Public Opinion Quarterly, 56,* 218–231.

Fuller, J. B., Patterson, C. E. P., Hester, K., & Stringer, S. Y. (1996). A quantitative review of research on charismatic leadership. *Psychological Reports, 78,* 271–287.

Ganster, D. C., Hennessey, H. W., & Luthans, F. (1983). Social desirability response effects: Three alternative models. *Academy of Management Journal, 26,* 321–331.

Gerstner, C. R., & Day, D. V. (1997). Meta-analytic review of leader–member exchange theory: Correlates and construct issues. *Journal of Applied Psychology, 82,* 827–844.

Gioia, D. A., & Sims, H. P., Jr. (1985). On avoiding the influence of implicit leadership theories in leader behavior descriptions. *Educational and Psychological Measurement, 45,* 217–232.

Greene, C. N., & Organ, D. W. (1973). An evaluation of causal models linking the received role with job satisfaction. *Administrative Science Quarterly, 18,* 95–103.

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.

Guzzo, R. A., Wagner, D. B., Maguire, E., Herr, B., & Hawley, C. (1986). Implicit theories and the evaluation of group process and performance. *Organizational Behavior and Human Decision Processes, 37,* 279–295.

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self–supervisor, self–peer, and peer–supervisor ratings. *Personnel Psychology, 41,* 43–62.

Harrison, D. A., & McLaughlin, M. E. (1993). Cognitive processes in self-report responses: Tests of item context effects in work attitude measures. *Journal of Applied Psychology, 78,* 129–140.

Harrison, D. A., McLaughlin, M. E., & Coalter, T. M. (1996). Context, cognition, and common method variance: Psychometric and verbal protocol evidence. *Organizational Behavior and Human Decision Processes, 68,* 246–261.

Harvey, R. J., Billings, R. S., & Nilan, K. J. (1985). Confirmatory factor analysis of the job diagnostic survey: Good news and bad news. *Journal of Applied Psychology, 70,* 461–468.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.

Hernandez, A., & Gonzalez-Roma, V. (2002). Analysis of multitrait–mulioccasion data: Additive versus multiplicative models. *Multivariate Behavioral Research, 37,* 59–87.

Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management, 21,* 967–988.

Idaszak, J., & Drasgow, F. (1987). A revision of the job diagnostic survey: Elimination of a measurement artifact. *Journal of Applied Psychology, 72,* 69–74.

Isen, A. M., & Baron, R. A. (1991). Positive affect as a factor in organi-

zational behavior. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 13, pp. 1–53). Greenwich, CT: JAI Press.

Iverson, R. D., & Maguire, C. (2000). The relationship between job and life satisfaction: Evidence from a remote mining community. *Human Relations, 53,* 807–839.

Jex, S. M., & Spector, P. E. (1996). The impact of negative affectivity on stressor strain relations: A replication and extension. *Work and Stress, 10,* 36–45.

Johns, G. (1994). How often were you absent—A review of the use of self-reported absence data. *Journal of Applied Psychology, 79,* 574–591.

Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Morristown, NJ: General Learning.

Judd, C., Drake, R., Downing, J., & Krosnick, J. (1991). Some dynamic properties of attitude structures: Context-induced response facilitation and polarization. *Journal of Personality and Social Psychology, 60,* 193–202.

Kemery, E. R., & Dunlap, W. P. (1986). Partialling factor scores does not control method variance: A rejoinder to Podsakoff and Todor. *Journal of Management, 12,* 525–530.

Kenny, D. A. (1979). *Correlation and causality.* New York: Wiley.

Kline, T. J. B., Sulsky, L. M., & Rever-Moriyama, S. D. (2000). Common method variance and specification errors: A practical approach to detection. *The Journal of Psychology, 134,* 401–421.

Korsgaard, M. A., & Roberson, L. (1995). Procedural justice in performance evaluation—The role of instrumental and noninstrumental voice in performance-appraisal discussions. *Journal of Management, 21,* 657–669.

Kumar, A., & Dillon, W. R. (1992). An integrative look at the use of additive and multiplicative covariance structure models in the analysis of MTMM data. *Journal of Marketing Research, 29,* 51–64.

Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait–correlated method and correlated uniqueness models for multitrait–multimethod data. *Psychological Methods, 7,* 228–244.

Law, K. S., & Wong, C. S. (1999). Multidimensional constructs in structural equation analysis: An illustration using the job perception and job satisfaction constructs. *Journal of Management, 25,* 143–160.

Lindell, M. K., & Brandt, C. J. (2000). Climate quality and climate consensus as mediators of the relationship between organizational antecedents and outcomes. *Journal of Applied Psychology, 85,* 331–348

Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional designs. *Journal of Applied Psychology, 86,* 114–121.

Lord, R. G., Binning, J. F., Rush, M. C., & Thomas, J. C. (1978). The effect of performance cues and leader behavior on questionnaire ratings of leadership behavior. *Organizational Behavior and Human Decision Processes, 21,* 27–39.

Lowe, K. B., Kroeck, K. G., & Sivasubramaniam, N. (1996). Effectiveness correlates of transformational and transactional leadership: A meta-analytic review of the MLQ literature. *Leadership Quarterly, 7,* 385–425.

MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin, 114,* 533–541.

MacKenzie, S. B., Podsakoff, P. M., & Fetter, R. (1991). Organizational citizenship behavior and objective productivity as determinants of managers' evaluations of performance. *Organizational Behavior and Human Decision Processes, 50,* 1–28.

MacKenzie, S. B., Podsakoff, P. M., & Fetter, R. (1993). The impact of organizational citizenship behaviors on evaluations of sales performance. *Journal of Marketing, 57,* 70–80.

MacKenzie, S. B., Podsakoff, P. M., & Paine, J. B. (1999). Do citizenship behaviors matter more for managers than for salespeople? *Journal of the Academy of Marketing Science, 27,* 396–410.

Marsh, H. W. (1989). Confirmatory factor analyses of multitrait–multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13,* 335–361.

Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analyses of multitrait–multimethod data: A comparison of alternative methods. *Applied Psychological Measurement, 15,* 47–70.

Marsh, H. W., & Yeung, A. S. (1999). The lability of psychological ratings: The chameleon effect in global self-esteem. *Personality and Social Psychology Bulletin, 25,* 49–64.

Martin, C. L., & Nagao, D. H. (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology, 74,* 72–80.

McGuire, W. J. (1966). The current status of cognitive consistency theories. In S. Feldman (Ed.), *Cognitive consistency* (pp. 1–46). New York: Academic Press.

Millsap, R. E. (1990). A cautionary note on the detection of method variance in multitrait–multimethod data. *Journal of Applied Psychology, 75,* 350–353.

Moorman, R. H., & Blakely, G. L. (1995). Individualism–collectivism as an individual difference predictor of organizational citizenship behavior. *Journal of Organizational Behavior, 16,* 127–142.

Mossholder, K. W., Bennett, N., Kemery, E. R., & Wesolowski, M. A. (1998). Relationships between bases of power and work reactions: The mediational role of procedural justice. *Journal of Management, 24,* 533–552.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15,* 263–280.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Olkin, I., & Finn, J. D. (1995). Correlations redux. *Psychological Bulletin, 118,* 155–164.

Organ, D. W., & Greene, C. N. (1981). The effects of formalization on professional involvement: A compensatory process approach. *Administrative Science Quarterly, 26,* 237–252.

Osgood, C. E., & Tannenbaum, P. H. (1955). The principle of congruity in the prediction of attitude change. *Psychological Review, 62,* 42–55.

Parker, C. P. (1999). A test of alternative hierarchical models of psychological climate: Pc_g, satisfaction, or common method variance? *Organizational Research Methods, 2,* 257–274.

Parkhe, A. (1993). Strategic alliance structuring: A game theoretic and transaction cost examination of interfirm cooperation. *Academy of Management Journal, 36,* 794–829.

Parrott, W. G., & Sabini, J. (1990). Mood and memory under natural conditions: Evidence for mood incongruent recall. *Journal of Personality and Social Psychology, 59,* 321–336.

Peterson, R. A. (2000). *Constructing effective questionnaires.* Thousand Oaks, CA: Sage.

Phillips, J. S., & Lord, R. G. (1986). Notes on the theoretical and practical consequences of implicit leadership theories for the future of leadership measurement. *Journal of Management, 12,* 31–41.

Podsakoff, P. M., & MacKenzie, S. B. (1994). Organizational citizenship behaviors and sales unit effectiveness. *Journal of Marketing Research, 31,* 351–363.

Podsakoff, P. M., MacKenzie, S. B., Moorman, R., & Fetter, R. (1990). The impact of transformational leader behaviors on employee trust, satisfaction, and organizational citizenship behaviors. *Leadership Quarterly, 1,* 107–142.

Podsakoff, P. M., MacKenzie, S. B., Paine, J. B., & Bachrach, D. G. (2000). Organizational citizenship behavior: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management, 26,* 513–563.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12,* 69–82.

Podsakoff, P. M., & Todor, W. D. (1985). Relationships between leader reward and punishment behavior and group processes and productivity. *Journal of Management, 11,* 55–73.

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology, 84,* 754–775.

Sackett, P. R., & Larson, J. R., Jr. (1990). Research strategies and tactics in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 419–489). Palo Alto, CA: Consulting Psychologists Press.

Salancik, G. R. (1982). Attitude–behavior consistencies as social logics. In M. P. Zanna, E. T. Higgins, & C. P. Herman (Eds.), *Consistency in social behavior: The Ontario Symposium, Vol. 2* (pp. 51–73). Hillsdale, NJ: Erlbaum.

Salancik, G. R. (1984). On priming, consistency, and order effects in job attitude assessment: With a note on current research. *Journal of Management, 10,* 250–254.

Salancik, G. R., & Pfeffer, J. (1977). An examination of the need–satisfaction models of job attitudes. *Administrative Science Quarterly, 22,* 427–456.

Schmitt, N. (1994). Method bias: The importance of theory and measurement. *Journal of Organizational Behavior, 15,* 393–398.

Schmitt, N., Nason, D. J., Whitney, D. J., & Pulakos, E. D. (1995). The impact of method effects on structural parameters in validation research. *Journal of Management, 21,* 159–174.

Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait–multimethod matrices. *Applied Psychological Measurement, 10,* 1–22.

Schriesheim, C. A. (1979). The similarity of individual-directed and group-directed leader behavior descriptions. *Academy of Management Journal, 22,* 345–355.

Schriesheim, C. A., Kinicki, A. J., & Schriesheim, J. F. (1979). The effect of leniency on leader behavior descriptions. *Organizational Behavior and Human Performance, 23,* 1–29.

Scullen, S. E. (1999). Using confirmatory factor analysis of correlated uniquenesses to estimate method variance in multitrait–multimethod matrices. *Organizational Research Methods, 2,* 275–292.

Shapiro, M. J. (1970). Discovering interviewer bias in open-ended survey responses. *Public Opinion Quarterly, 34,* 412–415.

Smither, J. W., Collins, H., & Buda, R. (1989). When rate satisfaction influences performance evaluations: A case of illusory correlation. *Journal of Applied Psychology, 74,* 599–605.

Spector, P. E. (1987). Method variance as an artifact in self-reported affect and perceptions at work: Myth or significant problem. *Journal of Applied Psychology, 72,* 438–443.

Spector, P. E. (1992). A consideration of the validity and meaning of self-report measures of job conditions. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 7, pp. 123–151). New York: Wiley.

Spector, P. E. (1994). Using self-report questionnaires in OB research: A comment on the use of a controversial method. *Journal of Organizational Behavior, 15,* 385–392.

Spector, P. E., & Brannick, M. T. (1995). A consideration of the validity and meaning of self-report measures of job conditions. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 10, pp. 249–274). New York: Wiley.

Spector, P. E., Chen, P. Y., & O'Connell, B. J. (2000). A longitudinal study of relations between job stressors and job strains while controlling for prior negative affectivity and strains. *Journal of Applied Psychology, 85,* 211–218.

Staw, B. M. (1975). Attribution of the "causes" of performance: A general alternative interpretation of cross-sectional research on organizations. *Organizational Behavior and Human Decision Processes, 13,* 414–432.

Stone, E. F. (1984). Misperceiving and/or misrepresenting the facts: A reply to Salancik. *Journal of Management, 10,* 255–258.

Stone, E. F., & Gueutal, H. G. (1984). On the premature death of need–satisfaction models: An investigation of Salancik and Pfeffer's views on priming and consistency artifacts. *Journal of Management, 10,* 237–249.

Strack, F., & Martin, L. (1987). Thinking, judging, and communicating: A process account of context effects in attitude surveys. In H. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 123–148). New York: Springer-Verlag.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology.* San Francisco: Jossey-Bass.

Thomas, K. W., & Kilmann, R. H. (1975). The social desirability variable in organizational research: An alternative explanation for reported findings. *Academy of Management Journal, 18,* 741–752.

Thurstone, L. (1927). A law of comparative judgment. *Psychological Review, 34,* 273–286.

Tourangeau, R., Rasinski, K., & D'Andrade, R. (1991). Attitude structure and belief accessibility. *Journal of Experimental Social Psychology, 27,* 48–75.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response.* Cambridge, England: Cambridge University Press.

Wagner, J. A., III, & Gooding, R. Z. (1987). Shared influence and organizational behavior: A meta-analysis of situational variables expected to moderate participation–outcome relationships. *Academy of Management Journal, 30,* 524–541.

Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Jounal of Educational Measurement, 24,* 185–201.

Watson, D., & Clark, L. A. (1984). Negative affectivity: The diposition to experience negative aversive emotional states. *Psychological Bulletin, 96,* 465–490.

Williams, L. J., & Anderson, S. E. (1994). An alternative approach to method effects by using latent-variable models: Applications in organizational behavior research. *Journal of Applied Psychology, 79,* 323–331.

Williams, L. J., & Brown, B. K. (1994). Method variance in organizational behavior and human resources research: Effects on correlations, path coefficients, and hypothesis testing. *Organizational Behavior and Human Decision Processes, 57,* 185–209.

Williams, L. J., Cote, J. A., & Buckley, M. R. (1989). Lack of method variance in self-reported affect and perceptions at work: Reality or artifact? *Journal of Applied Psychology, 74,* 462–468.

Williams, L. J., Gavin, M. B., & Williams, M. L. (1996). Measurement and nonmeasurement processes with negative affectivity and employee attitudes. *Journal of Applied Psychology, 81,* 88–101.

Winkler, J. D., Kanouse, D. E., & Ware, J. E., Jr. (1982). Controlling for acquiescence response set in scale development. *Journal of Applied Psychology, 67,* 555–561.

Wothke, W., & Browne, M. W. (1990). The direct product model for the MTMM matrix parameterized as a second order factor analysis model. *Psychometrika, 55,* 255–262.

Received June 6, 2002
Revision received February 10, 2003
Accepted February 18, 2003 ■