Sources of Method Bias in Social Science Research and Recommendations on How to Control It

Philip M. Podsakoff,[1] Scott B. MacKenzie,[2] and Nathan P. Podsakoff[3]

[1]Department of Management, Kelley School of Business, Indiana University, Bloomington, Indiana 47405; email: podsakof@indiana.edu

[2]Department of Marketing, Kelley School of Business, Indiana University, Bloomington, Indiana 47405; email: mackenz@indiana.edu

[3]Department of Management and Organizations, Eller College of Management, University of Arizona, Tucson, Arizona 85721; email: podsakof@email.arizona.edu

Annu. Rev. Psychol. 2012. 63:539–69

First published online as a Review in Advance on August 11, 2011

The *Annual Review of Psychology* is online at psych.annualreviews.org

This article's doi:
10.1146/annurev-psych-120710-100452

0066-4308/12/0110-0539$20.00

Keywords

common method variance, response style biases, marker variable technique, instrumental variable technique, unmeasured latent variable technique

Abstract

Despite the concern that has been expressed about potential method biases, and the pervasiveness of research settings with the potential to produce them, there is disagreement about whether they really are a problem for researchers in the behavioral sciences. Therefore, the purpose of this review is to explore the current state of knowledge about method biases. First, we explore the meaning of the terms "method" and "method bias" and then we examine whether method biases influence all measures equally. Next, we review the evidence of the effects that method biases have on individual measures and on the covariation between different constructs. Following this, we evaluate the procedural and statistical remedies that have been used to control method biases and provide recommendations for minimizing method bias.

Contents

INTRODUCTION 540
WHAT IS METHOD BIAS
AND WHY IS IT A PROBLEM? .. 541
What Is a Method? 541
What Is Method Bias? 542
Does Method Bias Affect All
Measures Equally? 543
EMPIRICAL EVIDENCE
OF THE EFFECTS OF
METHOD BIASES 543
Effects of General Method Bias on
Item Reliability or Validity 543
Effects of General Method Bias on
the Covariation Between
Constructs...................... 544
Estimates of Specific Types of
Method Bias on the Covariation
Between Constructs............. 545
WAYS TO CONTROL
FOR DIFFERENT SOURCES
OF METHOD BIAS 548
Procedural Remedies............... 548
Statistical Remedies 553
RECOMMENDATIONS 559
When Is Method Bias Likely
To Be a Problem? 559
What Can Be Done To Mitigate
the Problem? 562
CONCLUSION 565

INTRODUCTION

Over 50 years ago, Campbell & Fiske (1959) voiced their concerns about the biasing effects that methods of measurement may have on the validity of measures:

> In any given psychological measuring device, there are certain features or stimuli introduced specifically to represent the trait (construct) that it is intended to measure. There are other features which are characteristic of the method being employed, features which could also be present in efforts to measure quite different traits (constructs). The test, or rating scale, or other device, almost inevitably elicits systematic variance due to both groups of features. To the extent that irrelevant method variance contributes to the scores obtained, these scores are invalid. (Campbell & Fiske 1959, p. 84; words in parentheses added by present authors)

In the years since, a number of researchers have discussed a related problem—the biasing effects that measuring two or more constructs with the same method may have on estimates of the relationships between them. The major concern with measuring different constructs with the same method is the danger that at least some of the observed covariation between them may be due to the fact that they share the same method of measurement.

This concern with method bias is potentially important because the situations in which it is likely to be a problem are quite common. For example, Bodner (2006) reviewed the literature in six areas of psychology and found that most studies (76%) involved only a single measurement method, and of the studies that involved human subjects and adequately explained the measurement procedures, 33% involved self-report questionnaires as the sole measurement method. Similarly, Woszczynski & Whitman (2004) reviewed the studies reported in the top management information systems journals from 1996 to 2000 and found that 27% of the 428 articles published in this literature during this time period used a survey with self-reports as the predominant method of data gathering.

Unfortunately, despite the concern that has been expressed about method bias, and the pervasiveness of research settings with the potential to produce it, there is little agreement about whether it really is a problem for researchers. For example, although many authors believe that method bias is an important problem that needs to be controlled (e.g., Campbell & Fiske 1959; Cote & Buckley 1987, 1988; Doty & Glick 1998; Podsakoff et al. 2003; Podsakoff & Organ 1986; Sharma et al. 2009; Williams & Anderson 1994; Williams et al. 1989, 2010), some claim that it is a myth or urban legend (e.g., Chen & Spector 1991; Spector 1987, 2006; Spector & Brannick 2009).

Within the context of the above discussion, the purpose of this review is to explore the current state of knowledge about method biases. The first issue we explore is the question of what is method bias. It is obvious from reading the literature that scholars have different interpretations of what is meant by this term (e.g., Campbell & Fiske 1959, Edwards 2008, Lance et al. 2009, Messick 1991). In addition, there is a difference of opinion about what constitutes a bias (Cote & Buckley 1987, Spector & Brannick 2009, Williams et al. 1989). Is it the effect of method factors on the validity and reliability of individual measures, the covariation between measures of different constructs, or both? Finally, scholars also disagree about whether method bias affects all measures equally or some measures more than others (Lindell & Whitney 2001, Williams et al. 2010). Differences in these assumptions about the nature of method bias influence the way in which researchers try to control for it and the conclusions they reach regarding its effects.

The second issue we address is what the empirical evidence indicates about the extent to which method bias is a problem in behavioral research. For the purposes of this analysis, we examine the evidence of the effects of method factors on the reliability and validity of individual measures and the effects of method factors on the covariation between measures of different constructs.

The third issue we discuss is how researchers can control method biases. We start by reviewing the literature on the procedural and statistical remedies that researchers commonly use to control method biases, and we then discuss the strengths and limitations of each of these remedies for dealing with specific types of method biases.

In the final section, we address two related issues: (*a*) when method biases are likely to be a major problem in a study and (*b*) what researchers can do to mitigate their effects. We then conclude with a brief summary of the state of our knowledge about method biases in the behavioral sciences.

WHAT IS METHOD BIAS AND WHY IS IT A PROBLEM?

What Is a Method?

It is obvious from reading the literature that there are differences in how scholars define the term method. The term has traditionally been defined broadly to include several key aspects of the measurement process (Campbell & Fiske 1959, Fiske 1982). For example, according to Fiske (1982, p. 82),

> the term *method* encompasses potential influences at several levels of abstraction. Taking a paper-and-pencil instrument as an example, these influences include the content of the items, the response format, the general instructions and other features of the test-task as a whole, the characteristics of the examiner, other features of the total setting, and the reason why the subject is taking the test. Two units that have any one of these elements in common can show convergence due to that source, so the relationship obtained between them cannot safely be interpreted as associated with the traits or constructs in those units. For any single investigation, the only certain protection against this threat to validity is units using completely independent methods.

This is consistent with the views of most researchers (Bagozzi 1984, Baumgartner & Steenkamp 2001, Johnson et al. 2011, Messick 1991, Podsakoff et al. 2003, Siemsen et al. 2010, Weijters et al. 2010c), including Edwards (2008, p. 476), who argues that method biases arise from "response tendencies that raters apply across measures, similarities in item structure or wording that induce similar responses, the proximity of items in an instrument, and similarities in the medium, timing, or location in which measures are collected."

However, others have argued for a narrower definition (e.g., Lance et al. 2009, Sechrest et al. 2000). For example, Lance et al. (2009, p. 351) argue that the term method should be restricted to those measurement facets that represent "alternative approaches to assigning numbers to observations to represent [an individual's] standing on latent constructs." Based on this definition, Lance et al. (2009, 2010) include similarities in item content, structure, or format that induce similar responses and explicitly exclude effects due to response tendencies that raters apply across measures, occasions of measurement, and different situations in which measurement may occur. In addition, based on this definition we presume this also excludes item proximity and item order effects because these are not "alternative approaches to assigning numbers to observations."

For our part, we prefer the broader definition of method because regardless of whether one considers various rater response styles, item characteristics, and aspects of the measurement context to be "method" factors, they are all sources of systematic measurement error that threaten the validity of a study's findings. Indeed, if they are ignored they can threaten construct validity, distort the dimensional structure of psychological domains, and obscure relationships between constructs/traits (Messick 1991). Therefore, in the remainder of this review we adopt this broader conceptualization of the term method. In so doing we acknowledge Campbell & Fiske's (1959, p. 85) observation that, "The distinction between trait and method is of course relative to the test constructor's intent. What is an unwanted response set for one tester may be a trait for another who wishes to measure acquiescence, willingness to take an extreme stand, or tendency to attribute socially desirable attributes to oneself."

What Is Method Bias?

There is also disagreement about what constitutes a bias. Two detrimental effects produced by method factors have been recognized in the literature (e.g., Cote & Buckley 1987, 1988; Doty & Glick 1998; Podsakoff et al. 2003; Williams et al. 2010). The first detrimental effect is that method factors can bias estimates of construct reliability and validity (e.g., Bagozzi 1984, Baumgartner & Steenkamp 2001, Cote & Buckley 1987, Williams et al. 2010). A latent construct captures systematic variance among its measures. If systematic method variance is not controlled, this variance will be lumped together with systematic trait variance in the construct. This is a problem because it can lead to erroneous perceptions about the adequacy of a scale's reliability and convergent validity (Baumgartner & Steenkamp 2001, Lance in Brannick et al. 2010, Williams et al. 2010), and it can lead to underestimates of corrected correlations in meta-analyses because the reliability estimates will be artificially inflated due to method variance (Le et al. 2009).

In addition, Bollen (1989) demonstrated that in multiple regression models, uncontrolled systematic or random measurement error in a predictor can also bias estimates of the effects of other error-free predictors on a criterion variable even if this systematic measurement error is not shared with the criterion variable or with any of the other predictors. The direction of the bias will depend on the magnitude and sign of the relationships (*a*) between the imperfect predictor and the criterion variable and (*b*) between the imperfect predictor and the other predictors. Thus, although it is true that systematic measurement error in a predictor that is not shared with a criterion variable will tend to attenuate estimates of the effect of the predictor on the criterion variable (Spector & Brannick 2009), it can also bias estimates of the effects of other correlated predictors on the criterion variable.

The second important detrimental effect of uncontrolled method factors is that it can bias parameter estimates of the relationship between two different constructs. Several researchers (e.g., Baumgartner & Steenkamp 2001, Cote & Buckley 1988, Podsakoff et al. 2003, Siemsen et al. 2010) have demonstrated that method bias can inflate, deflate, or have no effect on estimates of the relationship between two constructs. Depending upon whether the method bias inflates or deflates the relationship, this is a serious problem because it can (*a*) affect hypothesis tests and lead to type I or type II errors, (*b*) lead to incorrect perceptions about how much variance is accounted for in a criterion construct, and (*c*) enhance or diminish the nomological or discriminant validity of a scale. Note, however, that Siemsen et al. (2010) and Evans (1985) have shown that although interaction and quadratic effects can be severely deflated by method bias, they cannot be artifacts of it.

It is for these reasons that, even though Spector & Brannick (2009, p. 348) argue that the effects of method factors on item validity and reliability are unimportant because they "do not speak to the issue of CMV and how it might inflate correlations," the overwhelming consensus among researchers is that both forms of bias are important and should be controlled whenever possible (Bagozzi & Yi 1990, Baumgartner & Steenkamp 2001, Cote & Buckley 1987, Doty & Glick 1998, Podsakoff et al. 2003, Siemsen et al. 2010, Williams et al. 2010).

Does Method Bias Affect All Measures Equally?

Finally, there is also some disagreement about how method bias affects the measures in a given study. Some researchers assume that if a method factor has any effect, it affects all measures equally. For example, researchers who use the correlational marker variable technique for controlling method bias (see table 1 in Williams et al. 2010) implicitly assume that a method factor has an equal effect on all measures because this technique is based on the assumption that, "the observed variables are contaminated by a single unmeasured factor that has an equal effect on all of them" (Lindell & Whitney 2001, p. 114). However, other researchers argue that method factors may have unequal effects on different measures. This is important because if equal effects are wrongly assumed when attempting to statistically control (or test) for method bias, the result will be the overestimation of the effect of method factors in some cases and the underestimation of them in others. Empirical tests of whether method factor loadings are equal or unequal have generally found support for the assumption of unequal effects of method bias (Rafferty & Griffin 2004, 2006; Williams et al. 2010). Similarly, Baumgartner & Steenkamp (2001) found that the proportion of variance in measures of different types of constructs that is attributable to specific response styles ranged from 0% to 29%. Finally, Cote & Buckley's (1987) meta-analytic estimates of the proportion of method variance in measures of different types of constructs ranged from 22% to 41%. Thus, the weight of the evidence suggests that method factors are likely to have unequal effects on different measures—whether they are different measures of the same construct (as in Rafferty & Griffin 2004, 2006; Williams et al. 2010) or measures of different constructs (as in Baumgartner & Steenkamp 2001, Cote & Buckley 1987).

EMPIRICAL EVIDENCE OF THE EFFECTS OF METHOD BIASES

Effects of General Method Bias on Item Reliability or Validity

Evidence of the impact of method biases on item validity and reliability comes from a number of meta-analyses of the results of confirmatory factor analyses of multi-trait multi-method (MTMM) matrices (e.g., Buckley et al. 1990, Cote & Buckley 1987, Doty & Glick 1998, Lance et al. 2010, Williams et al. 1989). These studies used previously published MTMM matrices to estimate confirmatory factor models with multiple trait and method factors. Typically, the correlations among the trait factors and among the method factors, but not between the trait and method factors, were estimated. A summary of these studies is provided in **Table 1**. Taken together, they indicate that 18% to 32% of the total variance in the items used in these studies was due to method factors.

Scherpenzeel & Saris (1997) went one step further by (*a*) estimating confirmatory factor models for 50 MTMM matrices involving

Table 1 Summary of studies using multi-trait multi-method matrices to partition trait and method variance in empirical relationships

Study	Sample	Variance attributable to trait factors	Variance attributable to method factors	Variance attributable to error
Cote & Buckley (1987)	70 matrices examining a wide variety of constructs	42%	26%	32%
Williams et al. (1989)[a]	11 matrices involving perceptions of jobs and work environments	48%	25%	21%
Buckley et al. (1990)	61 matrices examining a variety of constructs	42%	22%	36%
Doty & Glick (1998)	28 matrices	46%	32%	22%
Lance et al. (2010)	18 matrices	40%	18%	42%

[a]Values reported for variance estimates represent medians.

601 measures, (*b*) calculating the validity and reliability for each item, and then (*c*) examining the effect of 15 specific method factors on these item validities. Among the most important predictors of the item validities and reliabilities were the type of construct being measured, form and length of the response scale, social desirability of the item, mode of data collection, position of item in a battery of questions with the same instructions and response scale, and type of information requested (judgment, frequency, agree-disagree). Another study that examined the effect of specific types of method factors on item validity and reliability is that of Baumgartner & Steenkamp (2001). Across 60 measures of 11 constructs, they found that an average of 8% (ranging from 0% to 29%) of the variance in an item was due to five specific response sets/styles.

However, there are some potential criticisms of this MTMM-based evidence of method bias. One is that the estimates of the proportion of item variance due to method provided by these studies are not completely independent because the MTMM matrices they analyze overlap to some extent. However, the overlap is relatively small (about 13%). Another criticism is that trait and method variance becomes confounded as the correlations among traits and among methods increase. Specifically, Bagozzi (1993, p. 66) noted that when the correlations among the traits and among the methods are high, "the correlations among the method factors may represent the convergence of [a] general trait factor across methods, rather than true relationships among the methods." Whether the average method correlation of 0.47 found in the studies cited is large enough to support this interpretation is a matter of judgment. A third criticism is that serious problems resulting in nonconvergence and/or improper estimates can arise when attempting to fit a confirmatory factor model to MTMM data (Brannick & Spector 1990). However, the results of the studies reported in **Table 1** were based only on solutions that converged and had proper estimates. Thus, although these criticisms are important, we believe the MTMM-based evidence supports the general conclusion that method biases have an impact on individual item validities and reliabilities.

Effects of General Method Bias on the Covariation Between Constructs

Estimates based on MTMM meta-analytic studies. The results of MTMM studies can also be used to obtain estimates of the average effects of method biases on the correlation between different traits (or constructs). Assuming that trait, method, and random error interactions do not exist, Cote & Buckley (1988) show that the observed correlation between two variables x and y (R_{xy}) is equal to

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

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

More importantly, they demonstrate how Equation 1 and the variance estimates from MTMM meta-analytic studies can be used to decompose the average observed correlation between measures of two different traits that share the same method into the proportion due to (*a*) the correlation between the traits they represent and (*b*) the common method they share. For example, Cote & Buckley's (1987) meta-analysis reports that the average true correlation between traits across 70 MTMM samples was 0.674, the average percentage of trait variance in each measure was 0.417, the average true correlation between methods was 0.484, and the average percentage of method variance in each measure was 0.263. Therefore, using Equation 1, the average observed correlation can be decomposed into the proportion due to (*a*) the correlation between the traits they represent [$0.674^*\sqrt{0.417}^*\sqrt{0.417} = 0.281$] and (*b*) the common method they share [$0.484^*\sqrt{0.263}^*\sqrt{0.263} = 0.127$]. This suggests that the correlation between the traits was inflated approximately 45% (0.127/0.281) by method bias. Similar estimates of the percent of inflation due to method bias obtained from other meta-analyses of MTMM studies are 38% in Buckley et al. (1990), 92% in Doty & Glick (1998), and 60% in Lance et al. (2009).

Several points regarding these estimates are worth noting. First, these estimates are conservative because they are based on MTMM studies that used two or more less-than-perfectly correlated methods, and in many cases the biggest concern regarding method bias is in studies that use only a single method (which implies a true $R_{mk,ml}$ of 1.00). Indeed, if a single method had been used to calculate these estimates, they would have ranged from 94% to 270%. Second, the MTMM studies included in these meta-analyses overlap to some extent, so the estimates of inflation are not independent. Third, there is quite a bit of variance in the estimates—ranging from 38% to 92%. Nevertheless, regardless of which estimate is used, the bottom line is that the amount of method bias is substantial.

Estimates based on method-method pair meta-analytic technique. Another way to use meta-analytic data to estimate the impact of method biases has been proposed by Sharma et al. (2009). Their technique involves categorizing the meta-analytic correlations from previous studies on the basis of the susceptibility to method biases of the pair of methods used to measure the predictor and criterion. Their argument is that some method-method (*M-M*) pairs are more susceptible to method biases than others and that organizing the meta-analytic data on the basis of these pairings allows researchers to obtain an estimate of the effects that method biases have on the relationships of interest. To illustrate how this method can be applied, Sharma et al. (2009) conducted a meta-analysis of 75 samples of data reported in 48 studies examining the technology acceptance model. The results indicated that (*a*) the mean correlation between the focal constructs was about 0.16 when the susceptibility of the M-M pairs to method biases was low and about 0.59 when the susceptibility of the *M-M* pairs to method biases was high, and (*b*) about 56% of the between-studies variance in this literature was attributable to method biases. They concluded that method bias "presents a major potential validity threat to the findings of IS research" (Sharma et al. 2009, p. 474).

Estimates of Specific Types of Method Bias on the Covariation Between Constructs

The effects of same versus different sources. In addition to using meta-analytic techniques to assess the impact of method

Table 2 Summary of meta-analytic studies comparing same-source versus different-source relationships

	Estimates from same source				Estimates from different source				
Relationship	**k**	**N**	**r**	**ρ**	**k**	**N**	**r**	**ρ**	**% inflation**
Leader behaviors → outcome variables	255	2,874	0.414	0.456	255	2,354	0.156	0.191	239%
Personality variables → job performance	123	1,504	0.259	0.312	139	898	0.113	0.147	212%
Job attitudes → OCB	98	6,729	0.270	0.340	155	13,551	0.190	0.230	148%
Participative decision making → work outcomes	91	391	0.343	0.343	140	1,453	0.165	0.165	208%
Organizational commitment → job performance	148	3,745	0.180	0.183	159	1,924	0.138	0.138	133%
Person-organization fit → job performance	12	639	0.230	0.283	21	813	0.073	0.093	304%
OCB → performance evaluations	95	2,808	0.490	0.595	56	2,889	0.260	0.323	184%

Abbreviations: k, sum of number of studies used to calculate the averages across meta-analyses of a specific relationship; N, harmonic mean for the specific relationship using reported sample sizes across meta-analyses; OCB, organizational citizenship behavior; r, unweighted average of raw correlations reported in meta-analyses; ρ, unweighted average of corrected correlations reported in meta-analyses;% inflation was calculated as $(\rho_{\text{same source}}/\rho_{\text{different source}})$.

factors on responses to individual measures (e.g., Cote & Buckley 1987), these techniques have been used to assess the effects that method factors have on the strength of the relationships between two or more constructs. **Table 2** reports a summary of these types of meta-analytic studies. For this table, we searched for meta-analytic studies that had explored the moderating effect of the source of the ratings on the relationships that were examined, and we then combined the data across the different meta-analyses that had examined the same general content areas to get an estimate of the effects that same-source method biases had on the strength of the relationships reported. The results indicate that the average corrected correlation between leader behaviors and outcome variables (employee performance, ratings of leader effectiveness, etc.) when taken from the same source is 0.456, but only 0.191 when obtained from different sources. This means that the average corrected correlation between measures of leader behaviors and outcome variables is 239% (0.456/0.191) larger when these measures are obtained from the same source than when they are obtained from different sources. Similarly, the corrected correlation between measures of personality variables and job performance, job attitudes and organizational citizenship behaviors (OCBs), participative decision making and work outcomes, organizational commitment and job performance, person-organization fit and job performance, and OCB and performance evaluations are 213%, 147%, 208%, 133%, 304%, and 184% larger, respectively, when these measures are obtained from the same source than when they are obtained from different sources. Thus, it appears that the relationships between many widely studied constructs are strongly influenced by whether their measures are obtained from the same or different sources. However, it is important to recognize that although a large portion of the difference in the magnitudes of these correlations is undoubtedly due to method bias, some portion of it may also be due to the different perspectives of the raters on what constitutes job performance (Lance et al. 2010).

The effects of response styles. Using data from a large representative sample of consumers (N = 10,477) from 11 countries, Baumgartner & Steenkamp (2001) examined the biasing effects of acquiescence, disacquiescence, extreme, midpoint, and noncontingent response styles/sets on the correlations among 14 consumer constructs. Overall, they found that 27% of the variance in the magnitude of the 91 intercorrelations among the 14 constructs was due to the five response styles (64% of the variance was due to the traits). The effect of the response styles on the magnitude of the correlations among the constructs depended upon whether the (*a*) true correlation between the constructs was positive or negative and (*b*) response style components affecting the scales were positively or negatively correlated. If the true correlation between the constructs was positive and the correlation between the response styles was positive, they found that the magnitude of the observed correlation was inflated by 54% (the average correlation increased from 0.13 to 0.20). If the true correlation between the constructs was negative and the correlation between the response styles was negative, they found that the magnitude of the observed correlation was inflated by 67% (average correlation increased from −0.09 to −0.15). In contrast, if the true correlation between the constructs was positive and the correlation between the response styles was negative so that the substantive and response style components have opposing effects, they found that the observed correlation was deflated by 55% (average correlation decreased from 0.11 to 0.05); if the true correlation between the constructs was negative and the correlation between the response styles was positive, they found that the average observed correlation decreased from −0.07 to 0.01 and changed signs. They also found evidence that the amount of method bias varied across different types of constructs.

The effects of proximity and reversed items. Weijters et al. (2009) manipulated the proximity and the nature of the conceptual relationship between two items and examined their effects on the strength of the correlation between the items. Next, they specified a regression model that explained the correlation between all possible pairs of 76 items (N = 2,850) as a function of their distance apart on the questionnaire, and their conceptual relationship (nonreversed items measuring the same construct, reversed items measuring the same construct, or items measuring unrelated constructs). They found that, on average, two items measuring unrelated constructs had a correlation of only 0.04 when they were positioned six items apart, but the correlation increased to 0.09 when the items were positioned right next to each other. In other words, the correlation between unrelated items increases by 225% when they are positioned next to each other, as opposed to when they are positioned a few items further apart. For nonreversed item pairs, the average correlation significantly and substantially increases with decreasing interitem distance. When positioned six or more items apart, the average correlation between a pair of nonreversed items was 0.35, but this correlation increased to 0.62 (an increase of 177%) when these items were positioned next to each other. In contrast, when positioned six or more items apart, a reversed item pair had a correlation of −0.26, which decreased in magnitude to −0.06 when these item pairs were placed next to each other. Thus, Weijters et al. (2009, p. 7) concluded that up to a point, "correlations become weaker for nonreversed items and stronger for reversed items the further items are positioned from each other ..."

The effects of item wording. Harris & Bladen (1994) examined the effect of stress versus comfort item wording on the relationships between role ambiguity, role conflict, role overload, job satisfaction, and job tension. They found that the average correlation among these five constructs was 0.21 when item word bias was controlled but increased to 0.50 when it was not controlled (an increase of 238%). In addition, they found that the effect of method bias also varied depending upon the constructs involved.

The effects of item context. Harrison et al. (1996) manipulated the order of the questions measuring four constructs (voice, options, objectivity, and standards) to create either a positive or negative measurement context for the questions about outcome favorability and fairness perceptions. They found that the correlation between outcome favorability and fairness was only 0.10 in a positive measurement context but increased to 0.50 in a negative measurement context. This difference was significant and found to be due to the effect of the measurement context manipulation on the variance in fairness perceptions. They concluded that "researchers would have come to different substantive conclusions about the existence and strength of influences on fairness, solely because of the position in which proposed antecedents were measured" (Harrison et al. 1996, p. 257).

Taken together, the evidence presented in this section is not consistent with Spector & Brannick's (2009) assertion that "the effects of method have generally been small and rarely pose a threat" and instead supports Johnson et al.'s (2011) conclusion that "CMV is not an urban legend, but rather a specter that has the potential to haunt interpretations of observed relationships." Thus, it is no surprise that editors of major journals in several disciplines (Chang et al. 2010, Kozlowski 2009, Straub 2009, Zinkhan 2006) consider method biases an important problem that needs to be addressed.

WAYS TO CONTROL FOR DIFFERENT SOURCES OF METHOD BIAS

Procedural Remedies

Obtain measures of predictor and criterion variables from different sources. One obvious way to help control for method bias is to obtain the measures from different sources. There are two main ways this can be done: (*a*) obtain the predictor measure(s) from one person and the criterion measure(s) from another; or (*b*) obtain either the predictor or criterion measure(s) from one person and the other measure from secondary data sources (e.g., company records, annual reports). These procedures can diminish or eliminate the effects of consistency motifs, idiosyncratic implicit theories, social desirability tendencies, dispositional mood states, and tendencies on the part of the rater to acquiesce or respond in a lenient, moderate, or extreme manner because they make it impossible for the mindset of a common rater to bias the predictor-criterion relationship.

Evidence of the effectiveness of obtaining the predictor measure(s) from one person and the criterion measure(s) from another person is summarized in **Table 2**. The data reported in this table indicate that although the average corrected correlation between predictor and criterion variables was 0.359 when they were obtained from the same source, it decreased to 0.184 when they were obtained from different sources (a 49% decrease). Using a variation of this procedure, Ostroff et al. (2002) found that obtaining the predictor and criterion variables from different sources (rather than the same source) decreased the average split-level correlations between several dimensions of work climate and satisfaction by 71%. More specifically, their results indicated that separating the sources decreased the average split-level correlation from 0.07 to 0.02 when the individual was the unit of analysis, and from 0.24 to 0.07 when the department was the unit of analysis. Evidence of the effectiveness of controlling for method bias by obtaining either the predictor or criterion measure(s) from one person and the other measure from secondary data sources comes from two meta-analyses. First, a meta-analysis of research on the relationship between leadership style and effectiveness by Lowe et al. (1996) found that obtaining both the predictor and criterion variables from different sources decreased the correlation between leadership style and effectiveness by 67% (from 0.57 to 0.19) compared to when both measures were obtained from the same source. Second, a meta-analysis by Hulsheger et al. (2009) on the relationship between four team-process variables and team innovation found that obtaining both the predictor and criterion variables from different sources decreased the relationship by about 49% (from 0.45 to 0.23) compared to when both measures came from the same source.

Despite the fact that this approach seems to control for several important sources of method bias, it may not be appropriate to use in all cases. For example, this procedure is not appropriate when both the predictor and criterion variables are capturing an individual's perceptions, beliefs, judgments, or feelings. Beyond this, Chan (in Brannick et al. 2010) noted that this procedure is problematic for self-referential attitude and perception constructs because (*a*) the individual's perceptions may not translate into observable behaviors, (*b*) others may not have the opportunity to observe these behaviors, and (*c*) valid measurement by others requires them to accurately infer the individual's attitudes or perceptions based on the observation of the individual's behavior. Furthermore, this technique may not be feasible to use in all cases. For example, in some situations it may not be possible to obtain archival data that adequately represent one of the constructs of interest. In other situations, this technique may require more time, effort, and/or cost than the researcher can afford. In addition, when the sample size is small and the individual is the unit of analysis, the split-group procedure used by Ostroff et al. (2002) may not be feasible because it requires cutting the sample size in half and can result in too little power to detect the effects hypothesized.

Temporal, proximal, or psychological separation between predictor and criterion. Another way to control for method bias is to introduce a separation between the measures of the predictor and criterion variables (Feldman & Lynch 1988, Podsakoff et al. 2003). This separation may be (*a*) temporal (i.e., a time delay between measures is introduced), (*b*) proximal (i.e., the physical distance between measures is increased), or (*c*) psychological (i.e., a cover story is used to reduce the salience of the linkage between the predictor and criterion variables). Podsakoff et al. (2003) noted that these types of separation should reduce the respondent's ability and/or motivation to use previous answers to fill in gaps in what is recalled, infer missing details, or answer subsequent questions. A temporal separation does this by allowing previously recalled information to leave short-term memory, whereas a proximal separation does this by eliminating common retrieval cues, and a psychological separation does this by reducing the perceived relevance of the previously recalled information in short-term memory.

Evidence of the effectiveness of introducing a temporal separation between the measurement of the predictor and criterion variables comes from several studies. First, Ostroff et al. (2002) compared predictor-criterion variable correlations for concurrent ratings of both variables to ratings obtained after a one-hour or one-month delay. Although they found no significant differences in the average correlations between the concurrent and one-hour delay conditions, they reported that the average correlations were 32% lower after a one-month delay than they were in the concurrent condition. Second, Johnson et al. (2011, study 2) examined the effects of a three-week delay on the correlation between a latent predictor construct and a latent criterion construct. Their results indicated that the correlation between the constructs was 43% smaller after a three-week delay than it was when both were measured at the same time (although the design makes it unclear how much of this difference was due to the time delay rather than sample differences). Finally, in a firm-level analysis, Rindfleisch et al. (2008, table 5) found few significant differences in the correlations between several predictor and criterion variables after either no delay or a 30- to 36-month delay.

Although the weight of the evidence suggests that introducing a temporal separation is an effective means of controlling for some method biases, there are several disadvantages of this approach. First, introducing a temporal separation will obviously increase the complexity of the research design and potentially its cost. Second, when a temporal separation is introduced, it may allow other nonmethodological factors to influence the level of the outcome variable. Third, the longer the temporal delay, the greater the chance of respondent attrition. Fourth, it is difficult to determine what the appropriate delay should be for any given relationship, and it is likely that the appropriate delay varies across types of relationships. If the delay is too short, the temporal separation may be ineffective; and if the delay is too long, intervening factors are likely to affect the criterion variable. Fifth, and most importantly, the temporal separation procedure is based on the assumption that the true relationship between the constructs is relatively stable over the time period of the delay and that method bias will dissipate over time. If it is suspected that either of these assumptions is inaccurate, this method of control should not be used. Indeed, recent empirical research (Alessandri et al. 2010, Weijters et al. 2010a) suggests that the assumption that the method bias dissipates over time may be questionable.

Indirect evidence of the effectiveness of introducing a proximal separation between the measures of the predictor and criterion variables comes from studies demonstrating that separation attenuates method biases due to context effects (Tourangeau et al. 2000) and question order effects (Tourangeau et al. 2003). However, the most direct evidence comes from the study by Weijters et al. (2009) that found that proximal separation prevents the correlation between nonreversed (reversed) items from being artificially inflated (deflated). Based on their analysis, they concluded that researchers should try to position measures of the same-construct at least six items apart, separated by measures of other constructs using the same or different formats, or by means of dedicated buffer items.

Although there is some evidence that proximal separation is an effective means of controlling for some method biases, there are some disadvantages of this approach. First, it can increase the length of the questionnaire and that may cause fatigue, decrease response rates, or increase costs. Second, if the filler items are conceptually related to the measures of interest, they could create context effects that increase method bias.

Empirical evidence of the effectiveness of psychological separation as a means of reducing method bias is not readily available. However, there is no shortage of studies that recommend this procedure. For example, Aronson et al. (1998) note that one way to psychologically separate the predictor and criterion variables is to use a "multiple study" cover story, in which participants are told that for reasons of convenience or efficiency several unrelated studies are being conducted at the same time. This ruse is frequently employed in priming experiments (e.g., Higgins et al. 1977) and attitudinal research (e.g., Rosenberg 1965). Other ways to psychologically separate the predictor and criterion measures might be to (*a*) camouflage interest in the criterion or predictor variable by embedding it in the context of other questions so that it is less psychologically prominent (i.e., diminishing the salience of the measure) or (*b*) disguise the reasons for obtaining the predictor or criterion measure by leading respondents to believe that it is tangential to the main purpose of the study (i.e., making respondents think it is unimportant).

The principal disadvantage of this technique is that its effectiveness is dependent upon the credibility of the cover story, but a considerable amount of creativity and ingenuity is required to develop a convincing cover story. Consequently, it is essential to thoroughly pretest the cover story in order to ensure its effectiveness.

Eliminate common scale properties. Several authors (e.g., Campbell & Fiske 1959, Cronbach 1946, Feldman & Lynch 1988, Podsakoff et al. 2003, Tourangeau et al. 2000) have observed that method bias can result from common scale properties (i.e., scale type, number of scale points, anchor labels, polarity, etc.) shared by the items used to measure different constructs. For example, Feldman & Lynch (1988, p. 427) note that method bias "will occur to the extent that the question formats are perceived to be similar by respondents," because the similarity of the response format "enhances the probability that cognitions generated in answering one question will be retrieved to answer subsequent questions." The obvious remedy to this problem is to try to minimize the scale properties shared by the measures of the predictor and criterion variables.

Evidence of the effectiveness of this remedy comes from several studies. Kothandapani (1971) measured three constructs using four different scale formats (Likert, Thurstone, Guttman, and Guilford) and found that the average correlation was 0.45 when the criterion and predictor shared the same scale format and dropped to 0.18 when they did not share this scale property, a decrease of 60%. Arora (1982) measured three constructs using three different scale formats (Likert, semantic differential, and Stapel) and found that the average correlation was 0.34 when the criterion and predictor shared the same scale format and dropped to 0.23 when they did not share this scale property, a decrease of 32%. Flamer (1983) measured three constructs using three different scale formats (Likert, semantic differential, and Thurstone) in two different samples and found that the average correlation was 0.06 in sample A and 0.09 in sample B when the criterion and predictor shared the same scale format and dropped to 0.04 in sample A and 0.08 in sample B when they did not share this scale property, a decrease of 33% in sample A and 11% in sample B. Finally, Weijters et al. (2010c) examined the effect of common scale labeling on the correlation between attitudes and intentions and found that the average correlation was 0.69 when both the predictor and criterion variables had only the extreme end points of the scale labeled, and dropped to 0.60 when the criterion had the extreme end points labeled and the predictor had all points on the scale labeled, a decrease of 15%.

An advantage of this procedure is that it is often easy to translate some types of scale formats (e.g., Likert) into other formats (e.g., semantic differential) without changing the content of the item or other properties of the item (e.g., number of scale points). However, that is not always the case. For example, although it can be done, translating Likert or semantic differential items into Thurstone or Guttman scales is often difficult to do without altering the conceptual meaning of the measures (Nunally & Bernstein 1994). In these instances, it is important to give priority to maintaining the content validity of the items because a lack of content validity poses an even bigger threat to construct validity than does common method bias (MacKenzie et al. 2011). Thus, although minimizing common scale properties is always a good idea, there are practical limits to the extent to which this can be done.

Improving scale items to eliminate ambiguity. Ambiguous items are ones that are difficult to interpret and require people to construct their own idiosyncratic meanings for them. Johnson (2004) identifies several causes of item ambiguity, including the presence of indeterminate words such as "many" and "sometimes," words with multiple meanings, multiple ideas linked together with conjunctions or disjunctions, or complex constructions such as double negatives.

According to several authors (Cronbach 1950, Feldman & Lynch 1988, Podsakoff et al. 2003), the problem with ambiguous items is that they cause respondents to be uncertain about how to respond on the basis of the item's content, which increases the likelihood that their responses will be influenced by their systematic response tendencies (e.g., acquiescent, extreme, or midpoint response styles). The best solution to this problem is to make every effort to: keep questions simple, specific, and concise; define ambiguous or unfamiliar terms; decompose questions relating to more than one possibility into simpler, more focused questions; avoid vague concepts and provide examples when such concepts must be used; avoid double-barreled questions; and avoid complicated syntax (see Tourangeau et al. 2000). In addition, Krosnick (1991) notes that labeling every point on the response scale (rather than only the end points) is also an effective means of reducing item ambiguity. Unfortunately, we were unable to find any empirical evidence that specifically examined the effect of item ambiguity on estimates of the relationships between two different constructs.

Reducing social desirability bias in item wording. There is a great deal of evidence that items differ in perceived social desirability and that this affects responses to the item. For example, Edwards (1970) measured the impact of item social desirability on responses to the Minnesota Multiphasic Personality Inventory. He found that ratings of the social desirability of self-descriptive items made by a sample of judges were correlated 0.87 with endorsements of the items by another sample of subjects. This suggests that item wording can potentially undermine the accuracy of responses by causing subjects to edit their responses for social acceptability. If this editing affects both predictor and criterion measures in a similar manner, it could possibly bias the relationship between them. There are two commonly used procedures for controlling item social desirability (e.g., Kuncel & Tellegen 2009). The first is to obtain an assessment of the perceived social desirability of the items from judges and to revise the wording of the highly rated items to minimize or reduce the perceived level of social desirability. The other is to calculate the correlation between subjects' responses to each item and responses to a recognized social desirability scale (e.g., Paulhus 1984) and to revise the wording of items that correlate highly with this scale to minimize or reduce the perceived level of social desirability.

Although these procedures have been widely used and seem to have few disadvantages, we were unable to find any direct empirical evidence of their ability to prevent item social desirability from biasing the correlations between measures of different constructs. Moreover, implementing these procedures may be more difficult than it appears for two reasons. First, revising the items to eliminate their social desirability without compromising their content validity may be easier said than done. Second, Kuncel & Tellegen (2009, p. 201) have shown that "the relation between degree of endorsement of an item and its judged desirability level is often nonlinear and varies across items such that no general model of item desirability can be adopted that will accurately represent the relations across all items, traits, and trait levels." This would suggest that the linear correlation between responses to an item and responses to a social desirability scale may not always be a valid indication of the tendency of the item to evoke socially desirable responses.

Balancing positive and negative items. A number of authors (Baumgartner & Steenkamp 2001, Billiet & McClendon 2000, Mirowsky & Ross 1991, Weijters et al. 2010b) have noted that scale formats that ask respondents how strongly they agree or disagree with statements may be susceptible to acquiescence or disacquiescence response style biases. Respondents who exhibit acquiescence response styles tend to disproportionately use the positive side of the scale, whereas those that exhibit disacquiescence response styles tend to disproportionately use the negative side of the scale. As noted by Mirowsky & Ross (1991), these response style tendencies are problematic because they inflate the estimates of the reliability of measures, may produce misleading factor analytic solutions, and may inflate or deflate correlation and regression coefficients, depending on the type of questions that are asked. One procedural remedy that has been used to try to reduce this type of bias is "balancing" the positively worded (i.e., agreement with the item indicates a higher score on the underlying construct) and negatively worded (i.e., agreement with the item indicates a lower score on the underlying construct) measures of each construct. According to Baumgartner & Steenkamp (2001, p. 147), "although balanced scales do not eliminate the occurrence of acquiescence per se, they contain a built-in control for contamination of observed scores by yea-saying, because the bias is upward for half of the items and downward for the other half."

The advantage of this technique is that it is a proactive way to control for acquiescence and disacquiescence biases. However, there are several limitations of this technique (Baumgartner & Steenkamp 2001, Mirowsky & Ross 1991, Weijters et al. 2010b). First, many existing scales do not contain an equal number of positively and negatively worded items, and reversing the wording of some items may alter their content. Second, reversed items may be confusing for some respondents. Finally, empirical research suggests that because this technique does not always completely control for these biases, it should be used in conjunction with the statistical methods of control described in the next section.

Statistical Remedies

Although it is possible that the use of the procedural remedies discussed above will minimize the detrimental effects of method biases, researchers may not always be able to implement them beforehand. In these circumstances, they may find it useful to use one of the statistical remedies.

Unmeasured latent method factor technique. This is perhaps the oldest latent variable control technique (Bagozzi 1984, Bagozzi & Phillips 1982, Widaman 1985), and it has been used in approximately 50 studies (see Richardson et al. 2009). This technique involves adding a first-order method factor whose only measures are the indicators of the theoretical constructs of interest that share a common method. This technique has several advantages: (*a*) it does not require the researcher to measure the specific factor responsible for the method effect; (*b*) it models the effect of the method factor at the measurement level, rather than at the latent construct level (Schaubroeck et al. 1992, Williams et al. 1996); and (*c*) it does not require the effects of the method factor on each measure to be equal.

However, this approach has been criticized for several reasons. First, as noted by Podsakoff et al. (2003, p. 894), the unmeasured latent method 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." This is considered to be a serious flaw (e.g., Richardson et al. 2009), but it could also be considered a virtue, since it is desirable to control for all systematic sources of bias when testing hypotheses about the relations between constructs. Indeed, Phillips & Lord (1986) noted a similar confounding of method and substantive variance when trying to control for halo effects and concluded that there are advantages to controlling for both. Second, if the ratio of the number of indicators to the number of substantive constructs is low, the addition of a method factor can cause identification problems. Finally, this procedure is based on the assumption that the method factor does not interact with the trait factors; an assumption that has been questioned by several researchers (see Bagozzi & Yi 1990, Campbell & O'Connell 1967, Wothke & Browne 1990).

Correlation-based marker variable technique. This approach (Lindell & Whitney 2001) ideally requires researchers to (*a*) identify a "marker variable" that is expected for theoretical reasons to be completely unrelated to the substantive variables of interest, (*b*) use the smallest correlation between the marker variable and the substantive variables as an estimate of the effects of method bias, (*c*) adjust the zero-order correlation between every pair of substantive variables of interest by subtracting this estimate from the zero-order correlation between any pair of substantive variables and dividing by the quantity of 1 minus this estimate, and (*d*) examine whether the resulting partial correlation is significantly different from zero. They argue that if this partial correlation remains significant, the substantive relationships still hold even after controlling for method bias. According to Williams et al. (2010), this technique has been widely used in recent years.

The primary advantage of this technique is that it is easy to implement. The disadvantages are many. First, Lindell & Whitney (2001) do not require the marker variable to share any method characteristics with the substantive variables. Indeed, they suggest that the variable with the smallest correlation with the substantive variables can be arbitrarily selected in an ad hoc manner as a marker variable.

As noted by Williams et al. (2010, p. 507) this is problematic because if the marker variable does not share method characteristics with the substantive variables, "it cannot provide the vehicle for partialling out these biases from estimates of relations among substantive variables so as to obtain a 'truer' estimate of the relation, which is the goal behind the use of marker variables." Second, it assumes that method bias can only inflate and never deflate relationships among the substantive variables. Several researchers have demonstrated that this assumption is incorrect (Baumgartner & Steenkamp 2001, Cote & Buckley 1988, Podsakoff et al. 2003). Third, this technique ignores measurement error that could attenuate the correlations between the marker variable and the substantive variables that are used to obtain an estimate of method bias (Lance et al. 2010, Podsakoff et al. 2003, Williams et al. 2010). Fourth, this approach controls for method bias at the scale level rather than the item level (Williams et al. 2010). Fifth, this method is based on the assumption that the method factor represented by the marker variable does not interact with the substantive variables of interest, which has been disputed by several researchers (Bagozzi & Yi 1990, Campbell & O'Connell 1967, Wothke & Browne 1990). Sixth, it is based on the assumption that the smallest correlation between the marker variable and the substantive variables is a reasonable estimate of the effects of all types of method bias, which is not justified because the marker variable is not required to share any measurement characteristics (e.g., common scale format, anchors) with the substantive variables (cf. Podsakoff et al. 2003, Williams et al. 2010). Finally, this technique assumes that the method factor represented by the marker variable has an identical effect on every substantive variable of interest in the study. However, this assumption has been widely criticized (e.g., Podsakoff et al. 2003, Richardson et al. 2009, Sharma et al. 2009). Indeed, Williams et al. (2010) note that an analytical technique that can incorporate unequal method effect is needed in most organizational research settings because there is evidence that different types of variables contain differing amounts of method variance.

Regression-based marker variable technique. Siemsen et al. (2010) recently proposed that common method bias can be eliminated when estimating a regression equation subject to method bias by adding a marker variable that (*a*) is uncorrelated with the substantive variables of interest and (*b*) suffers from some type of method bias. In the event that the marker variable is modestly correlated with the substantive variables, their numerical analysis suggests that the addition of 3 to 5 variables is necessary.

Perhaps the greatest advantage of this technique is that it is easy to implement. However, there are several disadvantages. Like the correlational marker variable technique, this technique (*a*) ignores measurement error that could attenuate the correlation between the marker variable(s) and the substantive variables of interest, (*b*) controls for method bias at the scale level rather than the item level, and (*c*) is based on the assumption that the method factor represented by the marker variable(s) does not interact with the substantive variables of interest. In addition, this technique controls for only the net effect of the sources of method bias common to the marker variable(s) and the substantive variables and applies only to single-equation models. Furthermore, it is unclear what is being controlled by the addition of the marker variables. It is assumed to be "method bias" based on the subjective judgment that the marker variables are "theoretically unrelated" to the substantive variables. However, that may not be true. Finally, Siemsen et al. (2010, p. 472) limited their analysis to a single method factor even though they note that, "In practice, observed variables may suffer from multiple different methods factors . . . Although we expect our insights to hold if these methods factors are uncorrelated with each other, examining multiple correlated methods factors may lead to different results."

Instrumental variable technique. This technique is based on the fact that the presence of a method factor that influences both the predictor and the criterion variable in a model will cause the structural error term for the equation to be correlated with the predictor. In this instance, the supposedly exogenous predictor variable is really an endogenous predictor. This is an important problem because it violates an assumption of many estimation techniques [e.g., ordinary least squares (OLS), maximum likelihood (ML)] and causes the estimate of the effect of the predictor on the criterion variable to be biased (i.e., inconsistent). Antonakis et al. (2010) point out that method bias can be controlled, and an unbiased (i.e., consistent) estimate of the effect of the predictor on the criterion can be obtained, by adding appropriate instrumental variables (IVs) to the model and estimating the effect of the predictor on the criterion variable using two-stage least squares (2SLS). Briefly, in the first stage of the 2SLS estimation process the endogenous predictors are regressed on the IVs (and any other truly exogenous variables included in the model) to obtain predicted values for the endogenous predictors. In the second stage, the criterion variable is regressed on the predicted values of the endogenous predictors obtained in the first stage (and any other truly exogenous variables in the model). An instrumental variable is a truly exogenous variable (i.e., it does not depend on other variables) that is (*a*) correlated with the endogenous predictor for which it is to serve as an instrument and (*b*) uncorrelated with the structural error term for the equation. Thus, an IV is indirectly related to the criterion variable through the endogenous predictor but not directly related to the criterion variable. Antonakis et al. (2010) recommend adding at least one more IV than there are endogenous predictors in the model.

In order to be useful, each IV must satisfy two essential requirements (Antonakis et al. 2010, Kennedy 2008). First, the IV must be significantly and strongly related to the predictor it represents. This is required for two reasons. One is that IV estimators (although asymptotically unbiased) are biased in the same direction as OLS in small samples. Even in large samples, Kennedy (2008, p. 145) notes that the magnitude of this bias (*a*) can be quite large if the IV is not strongly correlated with the endogenous predictor it represents and (*b*) becomes even worse if several weak IVs are used. Indeed, even a slight correlation between a weak IV and the structural error term for the equation (perhaps caused by method bias) can cause the IV estimate to exhibit more bias (even asymptotically) than an OLS estimate. Another reason why strong IVs are required, according to Kennedy (2008, p. 145), is that "A weak instrument also causes the IV variance to be underestimated in small samples; this causes the true type I error rate to be higher than its chosen level." Thus, weak IVs lead to biased estimates and unreliable inference.

Second, the IV must be completely uncorrelated with the structural error term for the equation. Antonakis et al. (2010) emphasize that "the instruments must first pass a 'theoretical overidentification' test before an empirical one" because "if all the modeled instruments are not truly exogenous the overidentification test will not necessarily catch the misspecification." According to Kennedy (2008), this "theoretical overidentification test" for each IV should involve the use of existing literature and theory to (*a*) defend the implicit assumption that the IV is not an explanatory variable in the equation being estimated (i.e., that the IV does not have any direct effect on the criterion variable) and (*b*) explain why the IV could not be influenced by any of the method factors that influence the criterion variable or by any other omitted variables that affect the criterion variable (because if the IV is affected by these factors, it would be correlated with the structural error for the equation). Assuming that the IVs pass these theoretical tests, Antonakis et al. (2010) recommend using a Sargan chi-square test of overidentification to test empirically the assumption that the IVs are uncorrelated with the structural error term. If the IVs are unrelated to the structural error term, the overidentification tests will all be nonsignificant. If they are not, one must find better instruments.

The primary advantage of the instrumental variable technique is that it provides a straightforward solution to the problem of common method bias in situations where its causes cannot be identified or measured directly (Antonakis et al. 2010). However, perhaps the biggest disadvantage of this technique is the difficulty of selecting IVs that are related to the endogenous predictors and completely uncorrelated with the structural error term for the equation. Indeed, if (as is often the case) all of the possible sources of method bias that might affect the endogenous predictors and the criterion variable cannot be identified, then it is unclear how the IVs could pass Antonakis et al.'s theoretical overidentification test and how one could be confident that they were not affected by these unidentified method biases as well. In addition, it may prove to be difficult to identify IVs that are strongly related to the endogenous predictors, and as noted above, weak IVs lead to biased estimates and unreliable inference. Unfortunately, if these two requirements are not met, this technique will produce biased estimates and inflate the type I error rates, and researchers would be better off using another technique to control for method biases. A final disadvantage is that because the results are dependent upon the IVs selected, a test of the robustness of the second-stage estimates should probably be conducted.

CFA marker technique. To address some of the problems with the correlation-based marker variable technique, Williams et al. (2010) recommend using a series of marker variables that share measurement characteristics with the substantive variables of interest as indicators of a latent method factor. They propose a three-phase confirmatory factor analysis (CFA) marker technique to identify and control for method biases. Phase I of this analysis tests for the presence and quality of method effects associated with the latent marker variable. This phase requires specifying five different latent variable models and comparing their relative fit to each other. The first model (the CFA model) estimates loadings for each marker variable on a latent method factor and estimates all possible correlations among the method factor and the substantive constructs of interest, but it sets the loadings from the method factor to the indicators of the substantive constructs to zero. The second model requires that (*a*) the correlations between the method and substantive latent factors be set to zero, (*b*) the indicator loadings of the latent method factor be fixed at the estimates obtained from the CFA model, and (*c*) the loadings from the latent method factor to the indicators of the substantive constructs be set to zero. This model is called the baseline model because it serves as the baseline against which the method effects are assessed. The third model is called the method-C model (i.e., a constrained model). This model estimates the loadings from the latent method factor to the indicators of the substantive constructs but constrains these loadings to be equal to each other. A comparison of the fit of this model to the fit of the Baseline model provides a test of the assumption that the latent method factor has equal (tau equivalent) effects on the indicators of the substantive constructs of interest. In contrast to this assumption, the fourth model (called the method-U, or unconstrained, model) allows the loadings from the method factor to the indicators of the substantive variables to be freely estimated (i.e., unconstrained). A comparison of the fit of this model to the fit of the method-C model provides a test of the assumption that the method factor has unequal effects on the indicators of the substantive constructs. Finally, the fifth model, which is referred to as the method-R model (to represent restrictions on the parameters) is specified. This model is identical to the method-C and method-U models, with the exception that the correlations among the substantive constructs are constrained to the values estimated in the baseline model. A comparison of the fit of this model to the fit of either the method-C model or the method-U model (depending on which of these models fits the best) provides a test of the bias in the correlations among the substantive constructs that is due to the latent method factor.

Phase II of this analysis is devoted to quantifying how method variance affects the reliability of the substantive constructs. This is important because if method variance is not controlled, it will bias the reliability estimates of the substantive constructs. First, the completely standardized estimates of the factor loadings and error variances for each substantive construct from the baseline model are used to obtain reliability estimates for each construct (Werts et al. 1974). Next, the completely standardized substantive construct factor loadings, method factor loadings, and the error variances (from either the method-C or method-U model, depending upon which was supported) are used to decompose the total reliability calculated in the first step into the proportion due to the substantive construct and the proportion due to the method factor.

Finally, phase III is used to conduct a sensitivity analysis to increase confidence in the findings. Briefly, Williams et al. (2010) argue that since the amount of method variance associated with each indicator of the substantive constructs is represented by the magnitude of their loadings on the method factor, the sensitivity of the estimates of the correlations between the substantive constructs to method bias can be examined by substituting larger alternative values for these method factor loadings. The specific alternative values selected should be based on the confidence intervals of the unstandardized method factor loadings from either the method-C or method-U models (depending on which one was supported). The examination of this model allows researchers to determine the sensitivity of their results to increasing amounts of method variance associated with sampling error in the indicators.

Williams et al. (2010) have identified several advantages of this approach over the partial correlation approach proposed by Lindell & Whitney (2001). First, it models the effects of method biases at the indicator level (rather than construct level). Second, it provides a statistical test of method bias based on model comparisons. Third, it permits a test of whether method biases affect all measures equally or differentially.

Despite these advantages, there are a few potential problems regarding this approach that should be noted. First, it doesn't identify the nature of the method bias being controlled. Indeed, Williams et al. (2010, p. 507) note that, "without conceptual analysis of the nature of the marker variable, the meaning of its covariation with substantive variables cannot be understood." Related to this, a second problem is that the conceptual meaning of the latent method factor is ambiguous. Empirically, this construct is defined as the common variance among the marker variables. Although this technique requires the marker variables to be theoretically unrelated to the substantive constructs, it places no constraints on their theoretical relationships to each other. This means that potentially the marker variables could all come from a scale for a recognized construct (albeit one that is theoretically unrelated to the substantive constructs of interest). Consequently, it is unknown whether the common variance that empirically defines the marker variable construct is due to method artifacts or to some theoretically meaningful construct that is confounded with it. This would affect the loadings of the marker variables on this latent construct in Williams et al.'s CFA model as well as their baseline model (since it uses these estimates from the CFA model as fixed parameters).

Another problem is that the results are sensitive to the specific variables used as indicators of the latent method factor. This technique will only control for the net effect of the method characteristics that are shared by all of the marker variables and the indicators of the substantive constructs. If there are many relatively important method characteristics shared, this procedure will provide a strong test of method bias, but if there are only a few relatively unimportant method characteristics shared, this procedure will provide only a weak test of method bias.

A final problem is that in phases I and III, this procedure requires fixing parameter estimates in one model at specific values obtained from the estimation of an alternatively specified model. This two-step estimation process may not provide correct standard errors and goodness-of-fit statistics to test the fit of the resulting model (Kennedy 2008, Jöreskög 1998).

Directly measured latent method factor technique. To apply this technique, researchers must be able to anticipate the potential source of method bias and obtain measures of it. If direct measures of this particular source of method bias are available, bias can be controlled by adding to the theoretical model a method factor that has both the direct measures and the measures of the substantive constructs of interest as reflective indicators. This technique has been used in several studies (e.g., Bagozzi 1984, Schaubroeck et al. 1992, Williams & Anderson 1994, Williams et al. 1996). For example, Williams et al. (1996) used this technique to control for the effects of negative affectivity on the relationship between job attitudes and role perceptions. In general, it can be used to control for any contaminating factor for which direct measures are available (e.g., social desirability, positive affectivity).

The advantages of this approach are that (*a*) it unambiguously identifies the source of the method bias, (*b*) it controls for measurement error, (*c*) it models the effects of the biasing factor at the item level rather than at the construct level, and (*d*) it does not constrain the effects of the methods factor on the measures of the substantive construct to be equal. Perhaps the biggest disadvantage of this technique is that it requires researchers to anticipate the most important sources of method biases in their studies and to include measures of these sources. This is a serious problem because it is often difficult to identify the key sources of method bias in a given situation, and valid measures for these sources may not exist. In addition, this technique assumes that the method factor does not interact with the substantive constructs, which has been questioned by several researchers (Bagozzi & Yi 1990, Campbell & O'Connell 1967, Wothke & Browne 1990).

Measured response style technique. Another promising technique is to systematically measure common response styles and partial out their effects on responses. This procedure requires several steps. First, the relevant item population must be defined and a random sample taken of it to produce a representative heterogeneous set of items. As noted by Weijters et al. (2010b, p. 118), "The items should relate to constructs that do not form a meaningful nomological network." In order to develop reliable measures of the response styles, they recommend that a minimum of three sets of five items each should be used. Second, this random sample of heterogeneous items should be inserted as buffer items between the scales of substantive interest using the same scale format as for the other items on the questionnaire. Third, as many researchers have noted (e.g., Baumgartner & Steenkamp 2001; Weijters et al. 2008, 2010a,b), the most common response styles can be measured for each set of items as follows: (*a*) acquiescence response style (ARS)—calculate the extent of agreement with both positively and negatively worded items in each set (before negatively worded items have been reverse-scored), (*b*) disacquiescence response style (DRS)—calculate the extent of disagreement with both positively and negatively worded items in each set (before negatively worded items have been reverse-scored), (*c*) extreme response style (ERS)—calculate the proportion of items in each set on which the respondent endorses the most extreme (positive or negative) scale categories, and (*d*) midpoint response style (MRS)—calculate the proportion of items in each set on which the respondent endorses the middle scale category. Weijters et al. (2008, p. 414) provide an excellent illustration of how these operationalizations can be applied to a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). For each set of k items, they compute the measures of each response style as follows: ARS = $[f(5) \times 1 + f(6) \times 2 + f(7) \times 3]/k$; DRS = $[f(1) \times 3 + f(2) \times 2 + f(3) \times 1]/k$; ERS = $[f(1) + f(7)]/k$; MRS = $\mathrm{f}(4)/k$; where $f(o)$ refers to the frequency of response option o. This results in a measure of each of the four response styles for each of the sets of *k* items. Fourth, for each of the response styles, the measures obtained from each set of items are used as indicators of a latent construct. This means that, if there are three sets of items, there would be three indicators of each response-style latent construct.[1] Fifth, the latent constructs representing each response style are added to a latent variable model and their effects on the measures of the substantive constructs of interest are added.[2]

A few words of caution are in order. First, in order to ensure that the response-style measures only capture method variance, it is essential for the content of the set of items used to measure the response styles to be independent of the content of the measures of the substantive constructs. De Beuckelaer et al. (2010) found that using ad hoc sets of items is suboptimal at detecting ARS and ERS compared to using random heterogeneous items. Second, it is important to use a complete profile of response styles because it is difficult to decide a priori which response style may cause bias (Weijters et al. 2010a).

With these caveats in mind, if the effects of the response-style constructs on the measures of the substantive constructs are significant, it is evidence of method bias. However, if the estimate of the relationship between the constructs of interest is significant after controlling for these response styles, then one can be confident that the relationship is not solely due to these forms of method bias.

Using multiple indicators to measure response-style constructs has several advantages (Weijters et al. 2010b). First, it facilitates evaluation of the method construct(s) in terms of convergent and discriminant validity. Second, it allows for unique variances in the response-style item sets. Therefore, these unique variances are not confounded with the method construct(s) itself. Third, using multiple indicators of the method construct(s) enhances the stability of the model. Finally, unlike the unmeasured latent variable, marker variable, or CFA-marker variable approaches, it specifies the nature of the method construct (e.g., ARS, ERS) whose effects are being controlled. However, despite these advantages, there are also some limitations of this approach. First, it only controls for the response styles explicitly measured. Second, it requires the researchers to collect additional data to measure the response styles.

RECOMMENDATIONS

In this section, we suggest strategies for (*a*) identifying when method bias is likely to be a problem and (*b*) mitigating its effects. However, as Podsakoff et al. (2003, p. 899) emphasize, "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." The goal is to reduce the plausibility of method biases as a rival explanation for the relationships observed in a study.

When Is Method Bias Likely To Be a Problem?

There is widespread agreement that generating an optimal answer to even a single question can require a great deal of cognitive work, and the effort required to answer a long series of questions on a wide range of topics is substantial (Krosnick 1999, Sudman et al. 1996, Tourangeau et al. 2000). Although we may wish otherwise, not all respondents will be willing and able to exert the cognitive effort required

[1] Weijters et al. (2010b) found that the loadings of the response-style indicators on the method factors are essential tau equivalent (complemented with a time-invariant autoregressive effect): "This means that ARS and ERS are largely but not completely consistent over the course of a questionnaire" (p. 105). Note that they did not examine whether the method factor had tau equivalent effects on the measures of any other "substantive" constructs.

[2] When specifying this model, the covariances among the response styles should be estimated, and the indicators that are based on the same sets of items should have correlated error terms across response styles.

to generate accurate answers to the questions on a typical research instrument. What then? Krosnick (1999) argues that when the difficulty of the task of generating an optimal answer is high but a respondent's ability or motivation to expend the required amount of cognitive effort are low, respondents may "satisfice" rather than generate the most accurate answers by simply being less thorough in question comprehension, memory retrieval, judgment, and response selection. In our view, when respondents are satisficing rather than optimizing, they will be more likely to respond stylistically and their responses will be more susceptible to method bias. In other words, we expect that responses will be more strongly influenced by method bias when the respondents can't provide accurate responses (which is a function of their ability and the difficulty of the task) or when they are unwilling to try to provide accurate responses (which is a function of motivation).

Ability factors that may cause biased responding. The first question to consider is whether respondents are able to provide accurate answers, because if they are not, they may respond stylistically or be more susceptible to method bias. For example, Krosnick (1999) summarizes research that shows that respondents who are low in verbal ability or education are more likely to respond in a nondifferentiated manner when asked to rate objects on a single response scale (i.e., by giving all objects the same rating) and that nondifferentiated responding is more prevalent toward the end of a questionnaire due to fatigue. Similarly, there is evidence (e.g., Schwarz et al. 1992) that the amount of experience a respondent has had thinking about the topic of a question decreases his/her tendency to select the most recent of several response alternatives mentioned (regardless of content), presumably because it makes the respondent's knowledge of the topic more accessible.

Beyond this, Baumgartner & Steenkamp (2001) note that the tendency to agree with items regardless of content (i.e., ARS) can result from a respondent's low cognitive ability, poorly differentiated cognitive structure, or uncertainty about how to respond to the question. They also provide an excellent summary of several personality characteristics that are associated with biased responding. They cite research indicating that (*a*) "stimulation-seeking extroverts" may have a tendency to accept statements impulsively and agree with them regardless of content (i.e., ARS or positivity bias), (*b*) "controlled and reflective introverts who try to avoid external stimulation" may have a tendency to disagree with items regardless of content (i.e., DRS or negativity bias), and (*c*) respondents who are rigid, dogmatic, anxious, or intolerant of ambiguity may have a tendency to endorse the most extreme response categories regardless of content (i.e., ERS).

Thus, method biases and stylistic responding may be more likely to the extent that the respondents in a study have these ability limitations or possess these personality characteristics. Consequently, under these circumstances researchers would be wise to implement the appropriate procedural and statistical remedies discussed below.

Motivational factors that may cause biased responding. A second question to consider is whether respondents are motivated to provide accurate answers. Method biases and stylistic responding should be less likely to the extent that respondents are motivated to provide optimal responses to the questions and more likely to the extent that respondents are motivated to expend less effort by satisficing. Krosnick (1999) notes several factors that increase a respondent's motivation to exert the cognitive effort required to provide optimal answers including the need for cognition; the desire for self-expression, intellectual challenge, self-understanding, or emotional catharsis; and the desire to help employers improve working conditions, manufacturers produce better quality products, or governments make better-informed policy decisions. To the extent that respondents possess these needs/desires, they may be more likely to expend the effort required to generate an optimal answer, and the threat of method bias should be lower. In contrast, respondents may be motivated to minimize effort when they feel the questions are unimportant; believe their responses will not have useful consequences; feel compelled to participate in a survey to fulfill a course requirement; become fatigued by a seemingly unending stream of questions; or dislike the interviewer, experimenter, or source of the survey. To the extent that these things are true, respondents may be more likely to minimize their effort and rely on stylistic tendencies or other decision heuristics to arrive at a merely satisfactory answer.

In addition to considering the general factors that might motivate respondents to attempt to minimize effort by satisficing, researchers should also consider aspects of the measurement conditions that might increase the threat of specific types of bias. For example, researchers should consider the magnitude of the social consequences of a respondent's answers and the extent to which the measurement conditions make those consequences salient (see Paulhus 1984, Steenkamp et al. 2010). The more serious the social consequences of a particular response, the stronger a respondent's desire to provide a socially acceptable response is likely to be. Similarly, the more that the measurement conditions threaten a respondent's self-esteem, heighten his/her defensiveness, or increase the benefits (costs) of presenting a good (bad) impression, the more the respondent is likely to be motivated to respond in a socially desirable manner. Baumgartner & Steenkamp (2001) suggest that researchers should also consider whether aspects of the measurement context motivate respondents to conceal their true opinion by using the middle scale category regardless of their true feelings (MRS) or by responding to items carelessly, randomly, or nonpurposefully (NCR). The former may happen because respondents become suspicious about how their data will be used (Schmitt 1994), and the latter may happen because respondents are motivated to leave the testing situation, wish to rebel against a testing procedure, or are not motivated to invest the cognitive energy required to read and interpret questionnaire items (Jackson 1967). Finally, researchers should also consider whether respondents are likely to believe that two constructs are related by an implicit theory, because if they are, then the respondents may be motivated to provide answers that are consistent with that theory.

Task factors that may cause or facilitate biased responding. A third question that researchers should consider is the impact of the task on respondents. More specifically, researchers should evaluate the extent to which respondents will have difficulty generating accurate answers to the questions and the extent to which the measurement conditions may make it easy for them to minimize their effort by responding in a stylistic manner. For example, Doty & Glick (1998) argue that one reason why Cote & Buckley (1987) found that some types of measures contain more method variance than others is that responding to complex, abstract questions is a more difficult task for respondents than answering simple, concrete questions. In addition, they note that complex, abstract questions are more likely to trigger social psychological processes that increase the "covariation among the systematic error variance components, thereby increasing the bias in the observed relationships between constructs" (Doty & Glick 1998, p. 381).

Another task characteristic that makes it more difficult for respondents to provide accurate responses is item ambiguity. Because ambiguity makes respondents less certain about how to accurately answer a question (e.g., Podsakoff et al. 2003), it increases the likelihood that they will rely on their own stylistic response tendencies to generate a merely satisfactory answer and increases the sensitivity of their answers to context effects (see Tourangeau et al. 2000). Consequently, when evaluating the potential threat of method bias, researchers should consider the extent to which their questions fail to define ambiguous or unfamiliar terms, refer to vague concepts without providing clear examples, have complicated syntax, or are double-barreled. In addition, Krosnick (1991) notes that item ambiguity is greater if only the end points of a response scale are labeled (rather than every point).

In contrast, rather than making the task of providing an accurate response more difficult, other aspects of the measurement context may enhance the threat of method bias by making it easier to provide an alternative, merely satisfactory, response. For example, it is easier for respondents to provide answers that are consistent with each other or with an implicit theory if the answers to previous questions are readily available (physically or in memory) at the time of answering a later question. This is likely to be the case in a self-administered paper and pencil questionnaire and is often (but need not be) the case for online questionnaires. This may also be the case when questions are grouped together in close proximity by construct on the questionnaire. Alternatively, it seems plausible that ERS or MRS response styles would be easier to implement if the measures were grouped together by scale type, with the same number of scale points, with common anchor labels, and without any reversed item wording.

What Can Be Done To Mitigate the Problem?

Procedural remedies. Generally, studies should be designed to maximize respondent motivation and ability and minimize task difficulty so that respondents are more likely to respond accurately. To increase the probability that respondents can provide accurate answers to the questions, it is necessary to implement procedures that ensure that respondents have the ability to answer the questions asked, decrease the difficulty of responding accurately, and increase the difficulty of responding stylistically. To increase the probability that respondents will try to provide accurate answers, it is necessary to implement not only procedures that increase their motivation to provide accurate answers, but also procedures that decrease their motivation to respond stylistically by increasing the effort required to do so.

The key thing that must be done to make sure that respondents have the ability to answer questions accurately is to match the difficulty of the task of answering the questions with the capabilities of the respondents. One obvious way to do this is to make sure that you don't ask respondents to "tell more than they can know" (Ericsson & Simon 1980, Nisbett & Wilson 1977). This can be avoided by exercising caution when asking respondents about the motives for their behavior, the effects of situational factors on their behavior, or other things pertaining to cognitive processes that they are unlikely to have attended to or stored in short-term memory. Beyond this, researchers can decrease the difficulty of responding accurately by using clear and concise language, avoiding complicated syntax, defining ambiguous or unfamiliar terms, not referring to vague concepts without providing clear examples, avoiding double-barreled items, and labeling all scale points rather than just the end points.

Perhaps the easiest way to increase the probability that respondents will try to provide accurate answers to the questions is by developing a good cover story and instructions (Aronson et al. 1998). For example, the desire for self-expression or emotional catharsis may be enhanced by explaining in the cover story or instructions that "we value your opinion," "we need your feedback," or that we want respondents to "tell us what you think." The tendency to respond in a socially desirable manner, threats to self-esteem, and defensiveness may be diminished through anonymity, telling respondents in the cover story or instructions there are no right or wrong answers, and assuring them that people have different opinions about the issues addressed in the questionnaire. The motivation of respondents to provide accurate answers may also be increased by explaining how the information will be used or how it will benefit them or their organization (e.g., by mentioning that the data will help their employer to improve working conditions or make their job easier). Promising feedback to respondents may motivate them to respond more accurately so that they can gain greater self-understanding. Motivation can also be increased through endorsement of the study by senior management. Finally, motivation to respond accurately can be maintained by keeping the questionnaire short and minimizing redundancies to the extent possible. However, because multiple measures of the same construct are usually essential, the best approach may be to vary the wording of the items rather than just using synonyms.

In addition to increasing the motivation to respond accurately, it is also important to decrease the motivation to respond stylistically by increasing the effort required to do so. This can be done in several ways. The first is by reversing the wording of some of the items to balance the positively and negatively worded items. Of course, this is only a good idea if it can be done without altering the content validity or conceptual meaning of the scale and if the reverse-worded items are not confusing to respondents. A second way is by separating items on the questionnaire to eliminate proximity effects. However, this may not be feasible if the questionnaire is too short. A third way is by varying the scale types and anchor labels to the extent that it is conceptually appropriate.

Of course, the procedures outlined above are not likely to fully control for every type of method bias. For example, it is unlikely that self-deception biases, memory biases (e.g., things that were recently activated are more accessible), or perceptual biases (e.g., Gestalt principles of perception) would be controlled by these efforts. To the extent that these things are a concern, try to obtain the measures of the predictor and criterion constructs from different sources. This is most easily done if there are multiple observers of the phenomenon of interest who have access to the same information and if the phenomenon is not self-referential (Chan in Brannick et al. 2010). Under these circumstances, separating sources should help to diminish the effects of involuntary memory-based and perceptual biases and may help to reduce the biasing effects of stylistic responding. However, if these conditions are not met, this procedure may also introduce information biases or attribution biases.

If separating the sources is not feasible or desirable, another procedure that should help to diminish method bias is to separate the measurement of the predictor and criterion constructs temporally, methodologically, or psychologically. Temporal separation involves introducing a time lag between measurement of the predictor and criterion variables. This procedure is appropriate if (*a*) the phenomenon is not ephemeral, short lived, or rapidly changing; (*b*) the phenomenon is based on long-term (rather than short-term) memory effects; (*c*) a significant amount of respondent attrition is not likely to occur; and (*d*) it is financially and logistically feasible. To be effective, it is important for the temporal delay to be long enough to produce forgetting, clear short-term memory, or to disassociate cues in the two measurement occasions.

Methodological separation involves having respondents complete the measurement of the predictor variable under different methodological conditions than the criterion variables. For example, researchers can use different scale properties, response modes, and data collection locations for the predictor and criterion measures, or they can physically separate the predictor and criterion measures on the questionnaire. Methodological separation is appropriate provided that varying the scale properties or response mode does not alter the conceptual meaning of the measures and that the questionnaire is of sufficient length to separate the measures. This can diminish method bias by increasing the difficulty of responding stylistically, eliminating the saliency of any contextually provided retrieval cues, and/or reducing the respondent's ability to use previous answers to fill in gaps in what is recalled or to use prior responses to answer subsequent questions (Podsakoff et al. 2003).

The measures of the predictor and criterion variables can be separated psychologically by using a "multiple study" cover story, camouflaging interest in the criterion or predictor variable (i.e., by embedding it in the context of other questions so that it is less psychologically prominent), or disguising the reasons for obtaining the predictor or criterion measure (i.e., by leading respondents to believe that it is tangential to the main purpose of the study). These procedures diminish method biases by reducing the perceived diagnosticity of responses to the measures of the predictor variable as cues for how to respond to the measures of the criterion variable (cf. Feldman & Lynch 1988). However, psychological separation is unlikely to diminish biases due to the accessibility of responses to the measures of the predictor variable in memory. For this reason, it may be wise to use this procedure in conjunction with a temporal separation long enough to clear short-term memory. Of course, implementing this procedure (i.e., psychological separation) is contingent upon the researcher's ability to create a credible cover story, and it is only useful if the means of producing the psychological separation does not cause a temporal delay that is longer than the phenomenon of interest.

Statistical remedies. In situations where method bias is still an important concern, even after implementing procedural methods of control, we recommend that researchers follow this up with appropriate statistical remedies. More specifically, we recommend that researchers first try to use the directly measured latent factor technique or the measured response style technique because both of these techniques control for measurement error and specify the nature of the method bias. The former would be used if a researcher is concerned about a particular source of bias for which a valid measure of the biasing factor is available or could be developed. The latter would be used if a researcher is concerned with the biasing effects of response styles (e.g., ARS, ERS). In this case, we recommend following the guidelines outlined in Weijters et al. (2008).

If the specific source of the method bias is unknown or valid measures of the source of bias are not available, then we recommend using the CFA marker technique or the common method factor technique because these approaches control for measurement error, even though they do not clearly specify the nature of the method bias. The CFA marker technique requires the researcher to include appropriate marker variables that are theoretically unrelated to any of the measures of the focal constructs of interest in the questionnaire. The common method factor technique does not require the inclusion of any additional measures, but it is problematic because it may capture irrelevant trait variance in addition to systematic method variance.

A final technique that could be used to control statistically for method biases is the instrumental variable technique. Although it provides no insight into the nature of the method bias, if it could be properly implemented it would be effective. However, as we noted earlier, it is extremely difficult to identify instrumental variables that are strongly related to the endogenous predictor variables but completely uncorrelated with the structural error term for the equation. This is a serious barrier to implementing this technique because if these requirements are not met, this technique can produce biased estimates and inflate the type I error rates; even a slight correlation between a weak IV and the structural error term for the equation can cause the IV estimate to exhibit more bias (even asymptotically) than an OLS estimate (Kennedy 2008). Therefore, although the use of this technique to control for method bias is possible in principle, it may be difficult to put into practice.

Additional approaches. Two final approaches that might help rule out method bias as a rival explanation for a study's findings have been identified recently. The first alternative approach is based on the simulation findings of Evans (1985) and a proof by Siemsen et al. (2010), which demonstrate that although method bias can inflate (or deflate) bivariate linear relationships, it cannot inflate (but does deflate) quadratic and interaction effects. Consequently, if a study is designed to test hypotheses about quadratic or interaction effects, rather than main effects, then method bias would not be able to account for any statistically significant effects observed. Although this may not be possible or desirable in many instances, in those cases where it is conceptually appropriate and possible, it may be a reasonable alternative to the procedural and statistical remedies described above. The second alternative approach, recently suggested by Chan (in Brannick et al. 2010), is to (*a*) identify one or more potential sources of method bias, (*b*) manipulate them in the design of the study, and (*c*) test whether the hypothesized estimates of the relationships among the constructs generalize across conditions. Importantly, Chan notes that when used in combination with the statistical techniques described above, this method provides a powerful means of detecting and controlling method bias.

CONCLUSION

The purpose of this article has been to review the current state of knowledge about method biases. Our review indicates that although there is some disagreement about the way "method" and method "biases" are defined, the evidence shows that method biases can significantly influence item validities and reliabilities as well as the covariation between latent constructs. This suggests that researchers must be knowledgeable about the ways to control method biases that might be present in their studies. Consequently, we recommend procedural and statistical remedies that can be used to achieve this control. Although space constraints prevent us from addressing all of the issues regarding this important topic, we hope that we have provided some recommendations that researchers can use to deal with the detrimental effects of method biases in their research.

DISCLOSURE STATEMENT

The authors are unaware of any affiliation, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Alessandri G, Vecchione M, Fagnani C, Bentler PM, Barbaranelli C, et al. 2010. Much more than model fitting? Evidence for the heritability of method effect associated with positively worded items of the life orientation test revised. *Struct. Equ. Model.* 17:642–53

Antonakis J, Bendahan S, Jacquart P, Lalive R. 2010. On making causal claims: a review and recommendations. *Leadersh. Q.* 6:1086–20

Aronson E, Wilson TD, Brewer MB. 1998. Experimentation in social psychology. In *The Handbook of Social Psychology*, ed. DT Gilbert, ST Fiske, G Lindzey, Vol. 1, pp. 99–142. Boston, MA: McGraw-Hill. 4th ed.

Arora R. 1982. Validation of an S-O-R model for situation, enduring, and response components of involvement. *J. Mark. Res.* 19:505–16

Bagozzi RP. 1984. A prospectus for theory construction in marketing. *J. Mark.* 48:11–29

Bagozzi RP. 1993. Assessing construct-validity in personality research: applications to measures of self-esteem. *J. Res. Personal.* 27:49–87

Bagozzi RP, Phillips LW. 1982. Representing and testing organizational theories—a holistic construal. *Admin. Sci. Q.* 27:459–89

Bagozzi RP, Yi Y. 1990. Assessing method variance in multitrait-multimethod matrices: the case of self-reported affect and perceptions at work. *J. Appl. Psychol.* 75:547–60

Baumgartner H, Steenkamp JBEM. 2001. Response styles in marketing research: a cross-national investigation. *J. Mark. Res.* 38:143–56

Billiet JB, McClendon MJ. 2000. Modeling acquiescence in measurement models for two balanced sets of items. *Struct. Equ. Model.* 7:608–28

Bodner TE. 2006. Designs, participants, and measurement methods in psychological research. *Can. Psychol.* 47:263–72

Bollen KA. 1989. *Structural Equations with Latent Variables*. New York: Wiley

Brannick MT, Chan D, Conway JM, Lance CE, Spector PE. 2010. What is method variance and how can we cope with it? A panel discussion. *Organ. Res. Methods* 13:407–20

Brannick MT, Spector PE. 1990. Estimation problems in the block-diagonal model of the multitrait-multimethod matrix. *Appl. Psychol. Meas.* 14:325–39

Buckley MR, Cote JA, Comstock SM. 1990. Measurement errors in the behavioral sciences: the case of personality attitude research. *Educ. Psychol. Meas.* 50:447–74

Campbell DT, Fiske D. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* 56:81–105

Campbell DT, O'Connell EJ. 1967. Methods factors in multitrait-multimethod matrices: multiplicative rather than additive? *Multivar. Behav. Res.* 2:409–26

Chang S-J, van Wittleloostuijn A, Eden L. 2010. From the editors: common method variance in international business research. *J. Int. Bus. Stud.* 41:178–84

Chen PY, Spector PE. 1991. Negative affectivity as the underlying cause of correlations between stressors and strains. *J. Appl. Psychol.* 76:398–407

Cote JA, Buckley R. 1987. Estimating trait, method, and error variance: generalizing across 70 construct validation studies. *J. Mark. Res.* 24:315–18

Cote JA, Buckley R. 1988. Measurement error and theory testing in consumer research: an illustration of the importance of construct validation. *J. Consum. Res.* 14:579–82

Cronbach LJ. 1946. Response sets and test validity. *Educ. Psychol. Meas.* 6:475–94

Cronbach LJ. 1950. Further evidence on response sets and test validity. *Educ. Psychol. Meas.* 10:3–31

De Beuckelaer A, Weijters B, Rutten A. 2010. Using ad hoc measures for response styles: a cautionary note. *Qual. Quant.* 44:761–75

Doty DH, Glick WH. 1998. Common methods bias: Does common methods variance really bias results? *Organ. Res. Methods* 1:374–406

Edwards AL. 1970. *The Measurement of Personality Traits by Scales and Inventories*. New York: Holt, Rinehart & Winston

Edwards JR. 2008. To prosper, organizational psychology should … overcome methodological barriers to progress. *J. Organ. Behav.* 29:469–91

Ericsson KA, Simon HA. 1980. Verbal reports as data. *Psychol. Rev.* 87:215–57

Evans MG. 1985. A Monte Carlo study of the effects of correlated method variance in moderated multiple regression analysis. *Organ. Behav. Hum. Decis. Process.* 36:305–23

Feldman JM, Lynch JG. 1988. Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *J. Appl. Psychol.* 73:421–35

Fiske DW. 1982. Convergent-discriminant validation in measurements and research strategies. In *Forms of Validity in Research*, ed. D Brinbirg, LH Kidder, pp. 77–92. San Francisco, CA: Jossey-Bass

Flamer S. 1983. Assessment of the multitrait-multimethod matrix validity of Likert scales via confirmatory factor analysis. *Multivar. Behav. Res.* 18:275–308

Harris MM, Bladen A. 1994. Wording effects in the measurement of role conflict and role ambiguity: a multitrait-multimethod analysis. *J. Manage.* 20:887–901

Harrison DA, McLaughlin ME, Coalter TM. 1996. Context, cognition, and common method variance: psychometric and verbal protocol evidence. *Organ. Behav. Hum. Decis. Process.* 68:246–61

Higgins ET, Rholes WS, Jones CR. 1977. Category accessibility and impression formation. *J. Exp. Soc. Psychol.* 13:141–54

Hulsheger UR, Anderson N, Salgado JF. 2009. Team-level predictors of innovation at work: a comprehensive meta-analysis spanning three decades of research. *J. Appl. Psychol.* 94:1128–45

Jackson DN. 1967. Acquiescence response styles: problems of identification and control. In *Response Set in Personality Assessment*, ed. IA Berg, pp. 71–114. Chicago: Aldine

Johnson JA. 2004. The impact of item characteristics on item and scale validity. *Multivar. Behav. Res.* 39:273–302

Johnson RE, Rosen CC, Djurdevic E. 2011. Assessing the impact of common method variance on higher-order multidimensional constructs. *J. Appl. Psychol.* 96:744–61

Jöreskög KG. 1998. Interaction and nonlinear modeling: issues and approaches. In *Interaction and Nonlinear Effects in Structural Equation Modeling*, ed. RE Schumacker, GA Marcoulides, pp. 239–50. Mahwah, NJ: Erlbaum

Kennedy P. 2008. *A Guide to Econometrics*. Malden, MA: Blackwell. 6th ed.

Kozlowski S. 2009. Editorial. *J. Appl. Psychol.* 94:1–4

Kothandapani V. 1971. Validation of feeling, belief, and intention to act as three components of attitude and their contribution to prediction of contraceptive behavior. *J. Personal. Soc. Psychol.* 19:321–33

Krosnick JA. 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Appl. Cogn. Psychol.* 5:213–36

Krosnick JA. 1999. Survey research. *Annu. Rev. Psychol.* 50:537–67

Kuncel NR, Tellegen A. 2009. A conceptual and empirical reexamination of the measurement of the social desirability of items: implications for detecting desirable response style and scale development. *Pers. Psychol.* 62:201–228

Lance CE, Baranik LE, Lau AR, Scharlau EA. 2009. If it ain't trait it must be method: (mis)application of the multitrait-multimethod design in organizational research. In *Statistical and Methodological Myths and Urban Legends: Doctrine, Verity, and Fable in the Organizational and Social Sciences*, ed. CE Lance, RL Vandenberg, pp. 337–60. New York: Routledge

Lance CE, Dawson B, Birkelbach D, Hoffman BJ. 2010. Method effects, measurement error, and substantive conclusions. *Organ. Res. Methods* 13:407–20

Le H, Schmidt FL, Putka DJ. 2009. The multifaceted nature of measurement artifacts and its implications for estimating construct-level relationships. *Organ. Res. Methods* 12:165–200

Lindell MK, Whitney DJ. 2001. Accounting for common method variance in cross-sectional designs. *J. Appl. Psychol.* 86:114–21

Lowe KB, Kroeck KG, Sivasubramaniam N. 1996. Effectiveness correlates of transformational and transactional leadership: a meta-analytic review of the MLQ literature. *Leadersh. Q.* 7:385–425

MacKenzie SB, Podsakoff PM, Podsakoff NP. 2011. Construct measurement and validity assessment in behavioral research: integrating new and existing techniques. *MIS Q.* 35: 293–334

Messick S. 1991. Psychology and methodology of response styles. In *Improving the Inquiry in Social Science: A Volume in Honor of Lee J. Cronbach*, ed. RE Snow, DE Wiley, pp. 161–200. Hillsdale, NJ: Erlbaum

Mirowsky J, Ross CE. 1991. Eliminating defense and agreement bias from measures of the sense of control: a 2 × 2 index. *Soc. Psychol. Q.* 54:127–45

Nisbett RE, Wilson TD. 1977. Telling more than we can know: verbal reports on mental processes. *Psychol. Rev.* 84:231–59

Nunally JC, Bernstein IH. 1994. *Psychometric Theory*. New York: McGraw-Hill. 3rd ed.

Ostroff C, Kinicki AJ, Clark MA. 2002. Substantive and operational issues of response bias across levels of analysis: an example of climate-satisfaction relationships. *J. Appl. Psychol.* 87:355–68

Paulhus DL. 1984. Two-component models of socially desirable responding. *J. Personal. Soc. Psychol.* 46:598–609

Phillips JS, Lord RG. 1986. Notes on the theoretical and practical consequences of implicit leadership theories for the future of leadership measurement. *J. Manage.* 12:31–41

Podsakoff PM, MacKenzie SB, Lee J-Y, Podsakoff NP. 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J. Appl. Psychol.* 88:879–903

Podsakoff PM, Organ DW. 1986. Self-reports in organizational research—problems and prospects. *J. Manage.* 12:531–44

Rafferty AE, Griffin MA. 2004. Dimensions of transformational leadership: conceptual and empirical extensions. *Leadersh. Q.* 15:329–54

Rafferty AE, Griffin MA. 2006. Refining individualized consideration: distinguishing developmental leadership and supportive leadership. *J. Occup. Organ. Psychol.* 79:37–61

Richardson HA, Simmering MJ, Sturman MC. 2009. A tale of three perspectives: examining post hoc statistical techniques for detection and correction of common method variance. *Organ. Res. Methods* 12:762–800

Rindfleisch A, Malter AJ, Ganesan S, Moorman C. 2008. Cross-sectional versus longitudinal research: concepts, findings, and guidelines. *J. Mark. Res.* 45:261–79
Rosenberg MJ. 1965. When dissonance fails: on eliminating evaluation apprehension from attitude measurement. *J. Personal. Soc. Psychol.* 1:28–42
Scherpenzeel A, Saris W. 1997. The validity and reliability of survey questions: a meta-analysis of MTMM studies. *Soc. Methods Res.* 25:341–83
Schaubroeck J, Ganster DC, Fox ML. 1992. Dispositional affect and work-related stress. *J. Appl. Psychol.* 77:322–35
Schmitt N. 1994. Method bias: the importance of theory and measurement. *J. Organ. Behav.* 15:393–98
Schwarz N, Hippler HJ, Noelle-Neumann E. 1992. A cognitive model of response-order effects in survey measurement. In *Context Effects in Social and Psychological Research*, ed. N Schwarz, S Sudman, pp. 187–201. New York: Springer-Verlag
Sechrest L, Davis MF, Stickle TR, McKnight PE. 2000. Understanding method variance. In *Research Design: David Campbell's Legacy*, ed. L Bickman, Vol. 2, pp. 63–88. Thousand Oaks, CA: Sage
Sharma R, Yetton P, Crawford J. 2009. Estimating the effect of common method variance: the method-method pair technique with an illustration from TAM research. *MIS Q.* 33:473–90
Siemsen E, Roth A, Oliveira P. 2010. Common method bias in regression models with linear, quadratic, and interaction effects. *Organ. Res. Methods* 13:456–76
Spector PE. 1987. Method variance as an artifact in self-reported affect and perceptions at work: myth or significant problem. *J. Appl. Psychol.* 72:438–43
Spector PE. 2006. Method variance in organizational research: truth or urban legend? *Organ. Res. Methods* 9:221–32
Spector PE, Brannick MT. 2009. Common method variance or measurement bias? The problem and possible solutions. In *The Sage Handbook of Organizational Research Methods*, ed. DA Buchanan, A Bryman, pp. 346–62. Los Angeles, CA: Sage
Steenkamp JBEM, De Jong MG, Baumgartner H. 2010. Socially desirable response tendencies in survey research. *J. Mark. Res.* 47:199–214
Straub DW. 2009. Creating blue oceans of thought via highly citable articles. *MIS Q.* 33:iii–vii
Sudman S, Bradburn NM, Schwarz N. 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology.* San Francisco, CA: Jossey-Bass
Tourangeau R, Rips LJ, Rasinski KA. 2000. *The Psychology of Survey Response.* London: Cambridge Univ. Press
Tourangeau R, Singer E, Presser S. 2003. Context effects in attitude surveys—effects on remote items and impact on predictive validity. *Soc. Methods Res.* 31:486–513
Weijters B, Cabooter E, Schillewaert N. 2010c. The effect of rating scale format on response styles: the number of response categories and response category labels. *Int. J. Mark.* 27:236–47
Weijters B, Geuens M, Schillewaert N. 2009. The proximity effect: the role of inter-item distance on reverse-item bias. *Int. J. Res. Market.* 26:2–12
Weijters B, Geuens M, Schillewaert N. 2010a. The stability of individual response styles. *Psychol. Methods* 15:96–110
Weijters B, Geuens M, Schillewaert N. 2010b. The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Appl. Psychol. Meas.* 34:105–21
Weijters B, Schillewaert N, Geuens M. 2008. Assessing response styles across modes of data collection. *J. Acad. Mark. Sci.* 36:409–22
Werts CE, Linn RL, Joreskog KG. 1974. Intraclass reliability estimates—testing structural assumptions. *Educ. Psychol. Meas.* 34:25–33
Widaman KF. 1985. Hierarchically nested covariance structural models for multitrait-multimethod data. *Appl. Psychol. Meas.* 9:1–26
Williams LJ, Anderson SE. 1994. An alternative approach to method effects by using latent-variable models: applications in organizational behavior research. *J. Appl. Psychol.* 79:323–31
Williams LJ, Cote JA, Buckley MR. 1989. Lack of method variance in self-reported affect and perceptions at work: reality or artifact? *J. Appl. Psychol.* 74:462–68
Williams LJ, Gavin MB, Williams ML. 1996. Measurement and nonmeasurement processes with negative affectivity and employee attitudes. *J. Appl. Psychol.* 81:88–101

Williams LJ, Hartman N, Cavazotte F. 2010. Method variance and marker variables: a review and comprehensive CFA marker technique. *Organ. Res. Methods* 13:477–514

Woszczynski AB, Whitman ME. 2004. The problem of common method variance in IS research. In *Handbook of Information Systems Research*, ed. ME Whitman, AB Woszczynski, pp. 66–77. Hershey, PA: Idea Group

Wothke W, Browne MW. 1990. The direct product model for the MTMM matrix parameterized as a second order factor analysis model. *Psychometrika* 55:255–62

Zinkhan GM. 2006. Research traditions and patterns in marketing scholarship. *J. Acad. Mark. Sci.* 34:281–83