

Organizational Research Methods

<http://orm.sagepub.com>

Method Variance in Organizational Research: Truth or Urban Legend?

Paul E. Spector

Organizational Research Methods 2006; 9; 221

DOI: 10.1177/1094428105284955

The online version of this article can be found at:
<http://orm.sagepub.com/cgi/content/abstract/9/2/221>

Published by:



<http://www.sagepublications.com>

On behalf of:



The Research Methods Division of The Academy of Management

Additional services and information for *Organizational Research Methods* can be found at:

Email Alerts: <http://orm.sagepub.com/cgi/alerts>

Subscriptions: <http://orm.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations <http://orm.sagepub.com/cgi/content/refs/9/2/221>

Method Variance in Organizational Research

Truth or Urban Legend?

Paul E. Spector

University of South Florida

It has become widely accepted that correlations between variables measured with the same method, usually self-report surveys, are inflated due to the action of common method variance (CMV), despite a number of sources that suggest the problem is overstated. The author argues that the popular position suggesting CMV automatically affects variables measured with the same method is a distortion and oversimplification of the true state of affairs, reaching the status of urban legend. Empirical evidence is discussed casting doubt that the method itself produces systematic variance in observations that inflates correlations to any significant degree. It is suggested that the term *common method variance* be abandoned in favor of a focus on measurement bias that is the product of the interplay of constructs and methods by which they are assessed. A complex approach to dealing with potential biases involves their identification and control to rule them out as explanations for observed relationships using a variety of design strategies.

Keywords: *method variance; monomethod bias; measurement bias; construct validity*

It is quite widely believed that relationships between variables measured with the same method will be inflated due to the action of common method variance (CMV), also referred to as monomethod bias. Although a number of authors have noted that the CMV problem is overstated (e.g., Crampton & Wagner, 1994; Lindell & Whitney, 2001; Spector, 1987, 1994), statements suggesting that CMV is a serious problem persist, for example, "Most researchers agree that common method variance . . . is a potential problem in behavioral research" (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003, p. 879). Recent discussions of CMV note that its effect varies across measures of different constructs assessed with the same method, but it is still assumed that the source of CMV is the method itself. Furthermore, authors of empirical studies continue to try to explain away the possibility of CMV, whereas reviewers continue to suggest that monomethod studies are suspect and therefore unworthy of publication because of CMV.

The origin of the belief that CMV inflates correlations can be traced at least back to Campbell and Fiske (1959), who noted that a certain amount of variance in measurement can be attributable to the method used, citing as examples apparatus effects with Skinner boxes

Author's Note: An earlier version of this article was presented at the 2004 Academy of Management (AOM) conference in New Orleans. Thanks to Robert J. Vandenberg for organizing the Urban Legend AOM session and this special issue and for inviting me to participate and to Charles E. Lance for his helpful comments and suggestions on this article.

and formats with psychological tests. This suggests that there will be a method effect that will produce some degree of variance in all measures assessed with the same method. Because the method variance component would be shared or would be common across variables assessed with a given method, an inflation in relationships would occur over the relationships that should be observed due to the underlying theoretical constructs of interest. This idea is behind attempts to develop methods to estimate and/or adjust for the amount of CMV in measures that would produce a more accurate estimate of relationships among constructs (e.g., Lindell & Whitney, 2001; Podsakoff & Todor, 1985; see also Podsakoff et al., 2003, for a review of methods).

Interestingly, the concern for CMV seems to be raised almost exclusively when cross-sectional, self-report surveys are used. Monomethod studies (those using the same method for assessing all variables) using other approaches, such as reports about other people (e.g., assessment centers or job performance ratings), are less criticized for the same shortcoming, although some have noted that source bias can be a problem in these other domains (e.g., Lance, Baxter, & Mahan, 2005; Lance et al., 2000, although they question whether “source bias” is really due to method variance). This automatic criticism of the cross-sectional self-report has become invoked so broadly and often so automatically that I argue it has achieved the status of a methodological urban legend. The term *urban legend* is appropriate in that it reflects something that is based on truth but has been distorted and exaggerated as it is passed from person to person over time. We have all heard it so often and from so many sources that it does not occur to many of us to question the extent to which it is true.

Kernel of Truth: Method Affects Measurement

There are two fundamental truths when it comes to CMV. First, the way in which we measure something affects the numbers that are generated. For example, if we assess something that is socially sensitive with self-reports, such as a person’s level of negative affect, individuals who are high in social desirability (Chen, Dai, Spector, & Jex, 1997) will be likely to underreport their level of negative affect, showing that they are low even if they are not. Individuals who are low in social desirability are less likely to distort. This social desirability effect introduces systematic variance or bias into the assessment of the trait of interest. Not all constructs assessed via self-reports will be subject to social desirability (Moorman & Podsakoff, 1992), but they will likely be subject to some sorts of biases and errors associated with that method. Some we may be able to identify in terms of source, and some may be errors that defy explanation, for example, simple recording errors that might occur if a person is asked his or her age or gender. Furthermore, the assessment of some variables with a given method, such as trait anxiety with a self-report, are likely to be associated with more bias than others, such as age or gender.

The method of measurement can affect observed data in two ways: by changing the underlying construct of interest or by distorting the measurement process and not the construct itself. The former way is illustrated by the uncertainty principle from physics, suggesting that taking physical measurements can actually change the properties of an object, such as a nuclear particle. The same thing can happen with the assessment of people; for example, within a questionnaire, a person’s attitude might be affected by asking questions about

it, depending on the nature of the questions and how they are asked. Shadish, Cook, and Campbell (2002) listed reactivity of tests as a threat to the internal validity of a research design, noting how being assessed can have an impact on people's behavior. However, it seems unlikely that assessment would change all characteristics we might wish to measure; for example, we will hardly change a person's demographic characteristics, such as age or gender, by merely assessing them.

The latter way is of more concern in most cases and refers to potential biases that can affect measurement. Characteristics of instruments, people, and situations (see Podsakoff et al., 2003, for a review of these sources), as well as the nature of the construct, can allow bias into the measurement process. Some potential biases have been discussed in the literature, such as acquiescence response bias, negative affectivity, and social desirability. The latter two are considered personality variables in their own right, which might bias assessment of at least some self-reports.

Second, if the same method is used to assess two variables, if those two variables share a common source of bias, the correlation will likely be inflated, depending on how strongly related the two sources are. For example, if both trait anxiety and social adjustment are measured with self-reports, each might be biased by respondent social desirability, which will likely inflate observed correlations between anxiety and adjustment. However, if trait anxiety and gender are both assessed via self-report, it seems highly unlikely that reports of gender will be systematically distorted in the same way, and thus the correlation will be less likely to be inflated. In fact, under normal circumstances in which we conduct our studies, it seems unlikely that gender reports will be distorted at all. In this case, more likely the correlation will be attenuated because the high social desirability individuals who are also high in trait anxiety will underreport, thus introducing additional error into measurement that reduces the observed relationship with gender. Indeed, Williams and Brown (1994) showed how under many circumstances, method variance could attenuate correlations rather than inflate them.

Taken together, these two kernels of truths suggest that if we measure two or more variables with the same method, such as self-report, some of the observed correlations might be inflated due to shared biases. There are certainly cases in our literature in which observed variables shared sources of bias and may well have been inflated. However, just because some variables share biases does not mean that all variables share biases. The nature of shared bias depends on both the construct of interest and how it is measured, that is, the method (Spector & Brannick, 1995). The legend part is the assumption that method alone is sufficient to produce biases, so that everything measured with the same method shares some of the same biases, although some have suggested that the magnitude of CMV varies across measures using the same method. The reason this is legend is because there are few scientific data to unequivocally support this view and there are data to refute it.

Urban Legend: Everything Measured With the Same Method Shares CMV

If it is true that CMV is ubiquitous and everything measured via self-report or other single methods is plagued with it, we should be able to easily find supporting evidence. Certainly, there should be a number of specific biases, such as social desirability, that are widespread, leading to wholesale correlation inflation. Furthermore, a comparison of monomethod with

multimethod studies (those with different variables assessed with different methods) should find clear evidence of inflation whereby observed correlations are larger with monomethods than with multimethods. As I will show, such is not the case.

All Self-Report Measures Are Correlated

Perhaps the first piece of evidence that refutes the CMV legend can be easily found in many cross-sectional, self-report studies. If the self-report survey itself is a method that introduces shared bias into the measurement of variables, we should find a baseline level of correlation among all variables. Unless the strength of CMV is so small as to be inconsequential, this baseline should produce significant correlations among all variables reported in such studies, given there is sufficient power to detect them. Yet failure to find significant correlations, even those theoretically expected, is common in published studies that passed a peer-review process that disfavors null results. The work of Boswell, Boudreau, and Dunford (2004) is an example taken from a recent issue of the *Journal of Applied Psychology*. This study of the turnover process included 5 self-report variables from the same questionnaire assessing attitudes, motives, and perceptions. The sample size of 1,601 individuals was certainly large enough to detect even small amounts of CMV because correlations as small as .07 were statistically significant. Yet out of 10 correlations among the 5 self-report variables, 4 (40%) were nonsignificant, with the smallest being .02. Among the significant correlations, 3 were .10 or less, and the largest was only .20. This hardly supports the idea that CMV is a universal inflator of correlations. Counter to the CMV legend, using a self-report methodology is no guarantee of finding significant results, even with very large samples.

Potential Biasing Variables

Social desirability. There do exist variables that are considered potential biasing factors with self-reports, and these might be considered sources of CMV. As already noted, social desirability is one such variable studied in organizational research that might inflate observed correlations by producing CMV. Just how widespread its effects might be was addressed by Moorman and Podsakoff (1992) with two studies. First, they did an extensive literature search that uncovered 36 samples from 33 empirical studies linking a measure of social desirability to nine organizational variables. A meta-analysis of these studies found limited support for social desirability as a universal bias. The mean correlation across all 36 studies was only .05, with a range from .01 to .17 in magnitude across the different variables. Only three of the nine mean correlations exceeded .15 in magnitude, and four of the confidence intervals included zero, suggesting a nonsignificant relationship. So at best, social desirability accounts for a small amount of variance in a limited number of organizational variables. Of course, the variables most likely to relate to social desirability might not have been included in Moorman and Podsakoff's meta-analysis, but the argument is not that social desirability never influences correlations but rather that it often does not.

Second, Moorman and Podsakoff (1992) conducted a self-report study in which they included a measure of social desirability as well as a sample of organizational variables that included five of the variables from the meta-analysis that had confidence intervals not including zero. They reasoned that if social desirability was a bias that was inflating correlations, partialling it from correlations among biased variables should result in a correlation reduc-

tion. Thus, they compared zero-order correlations among their test variables to partials with social desirability controlled. Results found very little impact of social desirability, with most partials being within .02 of the zero orders. The largest difference was .04. Interestingly, in some cases, the partial correlations were larger than the zero orders, suggesting a suppressor effect.

Similarly, Ones, Viswesvaran, and Reiss (1996) conducted a meta-analysis to determine the possible biasing effects of social desirability on relationships between personality (Big Five) and several criteria including job performance and counterproductive work behavior. Like Moorman and Podsakoff (1992), they found variable and mostly small relationships of social desirability with the other variables in their study. The largest mean correlation they found was with emotional stability, which was .27. The remaining personality correlations ranged from 0 to .15. Correlations with performance criteria ranged from .19 for training performance (which was the only criterion variable with a confidence interval that did not include zero) to .00 for task performance. Furthermore, Ones et al. compared zero-order with SD-partialled correlations between personality and job performance and found no differences.

Overall, the results of both studies clearly suggest that social desirability is unlikely to have caused more than modest inflation of a few relationships and little or no inflation in most cases. Of course, there might well be other combinations of variables that would show larger effects, but my argument is not that social desirability cannot cause inflation but rather that its potential effects are limited to a fairly small subset of variables. Furthermore, even in those cases in which partial correlations are smaller than zero orders, there is no specific evidence for a biasing effect. In other words, it may be that those high in social desirability are accurately reporting their standing on constructs that in point of fact relate at the construct level to social desirability. Thus, evidence fails to support social desirability as a general source of correlation inflating CMV when self-reports are used.

Negative affectivity (NA). NA is another personality variable, like social desirability, that has been suggested as a source of bias that would produce CMV (Watson, Pennebaker, & Folger, 1987). These authors noted that individuals high in NA are predisposed to experience a variety of negative emotions that lead to a general negative view of the world. Thus, their self-reports are likely to be biased in a negative direction, leading them to report the job as stressful and dissatisfying. Watson et al. (1987) suggested that NA might act as a widespread biasing variable that inflates relationships among a large number of variables, particularly variables that reflect perceptions of job conditions and stressors, as well as strains such as job satisfaction and affective reactions. To the extent that reports of different variables are affected by this NA bias, correlations among them would be inflated.

The Watson et al. (1987) article led to a lively debate in the literature concerning the existence of NA bias (e.g., Brief, Burke, George, Robinson, & Webster, 1988; Chen & Spector, 1991). There are two questions that relate to the issue of method variance. First, does NA relate to organizational variables? And second, does NA have an impact on correlations among organizational variables? The answer to the first question is yes in that measures of NA relate to a variety of organizational measures. For example, Connolly and Viswesvaran (2000) conducted a meta-analysis showing a mean correlation of $-.27$ between NA and job satisfaction. Chen and Spector (1991) reported significant correlations of NA with a variety of self-reported job stressors and strains, such as role ambiguity, role conflict, interpersonal

conflict, situational constraints, frustration, anger, absenteeism, doctor visits, physiological symptoms, and intention of quitting.

The answer to the second question is more complex. Several sets of authors have reported comparisons of zero-order correlations to partial correlations with a measure of NA controlled (e.g., Brief et al., 1988; Chen & Spector, 1991). The extent to which zero-order correlations and partial correlations controlling for NA differ depends on the variables of interest. In some cases, differences are quite large, whereas for others, they are quite small. In fact, Frese (1985) even found that partial correlations for some variables were larger than zero-order correlations rather than smaller, suggesting a suppressor effect of NA.

Other authors have used structural equation modeling (SEM) to determine the effects of NA on structural models. For example, Chan (2001) investigated the impact of NA on relations among job satisfaction, perceived organizational support, organizational commitment, and intent to quit, controlling for NA. He found evidence that NA related to the organizational measures, but a comparison of models with and without NA included found little impact on relations among organizational variables. Similarly, Williams and Anderson (1994) compared structural equation models with and without NA, finding little impact of this potential bias on relations among job satisfaction, organizational commitment, leader contingent rewards, and job complexity.

If we were to accept that NA is a bias factor, the most reasonable conclusion based on existing data is that it affects only some variable combinations in ways that would have more than trivial effects on results. As with social desirability, there is no evidence for a universal effect; that is, NA is not invalidating all observed correlations among organizational variables, even those within the job stress domain that was the area of particular concern to Watson et al. (1987). However, even for those cases in which partial correlations were quite a bit smaller than zero orders, it is not clear that NA is acting as a bias. Spector, Zapf, Chen, and Frese (2000) provided evidence for several substantive mechanisms other than bias that might underlie the relationship of NA to other variables. In other words, the construct of NA might well relate to other constructs and not just bias their measurement.

One such possible mechanism is selection; that is, NA determines in part the nature of jobs people hold. This might be due to self-selection into jobs or to the differential ability of people to land jobs with favorable characteristics. In either case, this means that NA will correlate with some job-related variables. For example, Spector, Jex, and Chen (1995) found that high-NA individuals had jobs lower in autonomy and complexity as measured with job analysis techniques independent of the participants in the study. Spector, Fox, and Van Katwyk (1999) found that NA related to job complexity assessed by both job analysts and supervisors. Both studies provide evidence for a substantive rather than biasing effect of NA. Taken all together, there is no consistent evidence that NA is a constant source of CMV with self-reports that inflate correlations. Even in specific cases in which NA might bias correlations, there is conflicting evidence that questions whether it is really bias at all.

Acquiescence. A third potential source of bias is acquiescence: the tendency to agree with items independent of content. Certain individuals might tend to agree with everything, thus inflating relationships among self-report measures using agree-disagree response choices. Acquiescence could be a mechanism whereby test format as a method would produce CMV. The existence of acquiescence has been documented in the testing literature, and there is evidence that some individuals will exhibit this pattern of responding to items. However, the

idea that acquiescence acts as a shared bias across different measures was laid to rest by Rorer (1965), who demonstrated that although acquiescence might account for variance within a test, the acquiescence components were not common across tests. In other words, people who acquiesced differed across tests. Thus, acquiescence might act as error variance that would attenuate observed correlations to some extent rather than inflate them. Thus, it would seem that even within a specific test format, CMV is elusive.

Monomethod Versus Multimethod Correlations

If CMV is an epidemic that inflates correlations among self-report measures, one should expect that monomethod correlations will be larger than multimethod correlations. Indeed, there are many examples in which this is the case, but there are also many contrary examples as well.

One such case is Spector et al. (1995), who compared monomethod with multimethod correlations between personality and job characteristics. Incumbents completed questionnaires containing measures of trait anxiety, dispositional optimism, and job characteristics. Independent raters reviewed job descriptions and gave assessments of job characteristics. In 3 of 10 cases, the multimethod correlations were larger than the monomethod correlations. Crampton and Wagner (1994) conducted an ambitious analysis of more than 40,000 correlations from 581 articles to compare monomethod with multimethod correlations on the same variables. Out of 143 variable pairs they were able to compare, in 26.6% (38) cases, monomethod correlations were significantly higher than multimethod correlations, in 11.2% (16) they were lower, and in 62.2% (89) of the cases, there was no significant difference. As Crampton and Wagner concluded, these results fail to support the idea that CMV is a universal problem, but rather it was a concern with only some combinations of variables.

Doty and Glick (1998) conducted SEM and meta-analysis on data from 28 multisource studies. They estimated that on average (median), method variance biased observed correlations among the underlying constructs by 26%, which they concluded did not necessarily invalidate many conclusions based on monomethod results. There are two things to keep in mind about these sorts of analyses. First, inferences about method variance are being drawn based on a comparison of monomethod versus multimethod correlations on measures of the same constructs. This assumes that the correlations based on mixed-methods data are more accurate than correlations based on single-method data, so differences are due to method. However, it is possible that the methods are not equally valid measures of the underlying constructs (Frese & Zapf, 1988). Thus, correlations from data that mixed methods might be underestimates of true relationships, making estimates of method variance inflated. Second, even if we accept these estimates as accurate reflections of CMV inflation, there was considerable variation from study to study. Again, both the method and traits matter, and one cannot conclude that method alone (i.e., self-report) is producing CMV.

What Is a Method?

An important issue we quickly confront when dealing with CMV concerns what we mean by a method. Campbell and Fiske (1959) noted that different item formats within a questionnaire could be considered different methods in that there can be CMV attributable to each format. However, methods can vary in a large number of ways, and it is not always clear what the

critical features of methods might be that can define them. Such clarity is necessary to devise effective research strategies that allow for confident inference.

Doty and Glick (1988) discussed the nature of methods, providing a taxonomy that classified measures along a facet of measurement technique versus data source. Measurement technique differences involve item formats, item wording, and data collection procedures, such as questionnaire versus interview. Data source is whether data come from a single rater (incumbent) or multiple raters (e.g., incumbent, coworker, or supervisor). Although this taxonomy provides a good start, more work needs to be done in linking features of methods with specific sources of variance. For example, forced-choice formats were developed specifically to control for social desirability. Coworker or supervisor ratings might be helpful in controlling for self-serving biases, such as with self-appraisals of performance. However, these methods might control for some forms of bias but not others, and they might introduce other biases or problems. For example, nonincumbent ratings of incumbent job characteristics have been shown to have less discriminant validity than incumbents' own ratings (e.g., Glick, Jenkins, & Gupta, 1986). So the multisource design is not without its own limitations.

The literature on generalizability theory is also relevant to the issue of method (Shavelson & Webb, 1991), providing procedures for partitioning variance in measurement according to different facets that describe the conditions and method of measurement. For example, facets include the targets being assessed (e.g., people), items, occasions, raters, and scales (Shavelson & Webb, 1991), all of which can be considered aspects of method.

How Should We Deal With CMV?

Perhaps the first step in dealing with the potential problem of CMV is to change our thinking about it. CMV as originally discussed by Campbell and Fiske (1959) is variance due to the use of a specific method regardless of the construct of interest. More recent discussions (e.g., Podsakoff et al., 2003) have considered CMV as something that can affect different constructs assessed with a given method to varying degrees. In factor analytic terminology, different constructs would have different sized loadings on a method factor. This still assumes that CMV exists but that its effect can be variable. My argument is that CMV is an urban legend, and the time has come to retire the idea and the term, replacing it with a more complex conception of the connection between constructs and their assessment. Rather than accepting the idea that there is systematic variance produced by a particular method, we should instead think for each measured variable what the likely sources of variance might be and how different features of method might control them. For example, if our interest is in variables shown to relate to social desirability, we might consider using methods that would control it, such as relying on observer, peer, or supervisor ratings. Of course, there is no guarantee that individuals' social desirability will not affect their observable, public behavior if they know they are being observed and introduce bias in the observer's assessment. If a concern, on the other hand, is the effect of mood, a different strategy might be chosen, such as separating the measurement of variables over time.

Part of the design of our study should involve a careful analysis of our purpose and the nature of our desired inference in relation to the measurement methods we will use. Are we interested in how perceptions of justice relate to job attitudes, or are we interested in how the objective work environment leads to justice perceptions? The first question might be reasonably addressed with monomethod self-reports, but the latter will require a more complex

approach. If self-reports are chosen, we should be careful about what we can expect subjects to accurately report, what sorts of biases might be introduced, and what sorts of conclusions are most reasonable. Certainly, we expect that people are able to report many internal states, including attitudes, emotions, perceptions, and values. However, people might not be able to report accurately on the objective environment, depending on the nature of what we are asking them to tell us. Factual information, such as whether they have a private office at work or their age, might be relatively impervious to most biases. More abstract social constructs such as autonomy or role ambiguity will introduce a certain level of subjectivity that leaves room for a variety of biases.

There are a number of design and measurement strategies that can be helpful in controlling for and ruling out biases. The effectiveness of a given strategy is dependent on the nature of the construct of interest and the means of assessing it. A large part of the problem is that viable alternative methods are not always available. For example, it is difficult to get accurate information about internal states, such as attitudes or emotions, with anything other than self-reports. The trick is to minimize possible biases through the design of measures or to link self-reports to measures using other methods that would provide confirmation about an observed relationship between variables. If the alternative method is less accurate, which is likely with internal psychological states, finding smaller relationships should not be automatically attributed to method variance inflation within the single method. It is equally likely that the multimethod relationship was attenuated.

Time can be an effective means of controlling occasion factors that influence measurement at a given point in time. For example, a person's mood at the time they complete a questionnaire can affect responses to some questions. Assessing different variables on different occasions can help reduce such biases, but there are two complications. First, one must know the time frame of the occasion factor; for example, how long does it take for mood to change? Second, one must be careful that the occasion factor acted as a bias in affecting assessment and did not affect the underlying construct itself. If the latter, observed correlations over time might not be accurate.

Nonincumbent raters, such as observers, peers, or supervisors, are often used to minimize potential biases that might be inherent to monomethod studies relying on all self-report. Such methods can be used to control self-serving biases, social desirability, and other possible within-subject factors that might distort correlations among variables of interest. As noted earlier, often such alternative sources are inaccurate (Frese & Zapf, 1988), and they often suffer from poor discriminant validity (e.g., Glick et al., 1986). Furthermore, they cannot control for all biases, as the incumbent and alternative source might share a bias, especially if there is contact between them. For example, incumbent mood might serve as a third variable even for an observer who might be influenced by the incumbent's apparent and observable emotional expression. The same workplace incident that biased an incumbent might bias a coworker or the supervisor as well.

Objective measures are sometimes available that might be resistant to many of the biases that can distort human judgments and reports. Although such measures are valuable, they are not entirely without bias. For example, absence is a factual and verifiable event. However, records are not always accurate, and errors might well be systematic. For example, individuals who are high (or low) on certain characteristics might be more (or less) likely to fail to report sick leave. Nevertheless, objective measures can be quite useful when available in controlling many biases.

Statistical control can be used to rule out plausible biases as long as those biases can be assessed. For example, the possible biasing effect of negative affectivity can be explored by including an NA measure and comparing results with and without NA controlled. Of course, there is an asymmetry between ruling in and ruling out the effect of a bias. If controlling the bias factor has little impact on observed relationships, we can be fairly confident that particular variable was not a problem in the study. That does not mean, of course, that there were not other unmeasured biases. However, finding a reduction in relationship after control is inconclusive. It might be that the potential bias was distorting results, but it also might mean that the potential bias played a substantive role. For example, perhaps NA was a cause or effect of both variables of interest. The observation that entering a control affected results supports the possibility of bias, but it is itself very weak evidence.

If CMV were not an urban legend but rather each method produced a certain amount of method variance, it might be relatively easy to use statistical methods to estimate and control it. Unfortunately, the methods that exist to estimate and control CMV (see Podsakoff et al., 2003, for a review of them) have limitations and in many cases are controlling for something that does not exist. For example, researchers who compare methods often assume that higher monomethod than heteromethod correlations should be attributed to CMV, but that is not always a safe assumption. For many constructs, an incumbent will be a more valid source of data than an alternative source, rendering the all-self-report study more accurate than one mixing incumbent with an alternative source. In fact, Frese and Zapf (1988) discussed how correlations crossing different sources tended to underestimate relationships among constructs in many cases.

A more useful approach is to assume each operationalization of a variable (or method-trait combination) carries with it a unique set of potential biases, and operationalizations of different variables can share biases. What is necessary is identifying through both conceptual and empirical work potential biasing factors and then researching possible effects of such factors on observed relationships among variables of interest. Some design and statistical strategies can be useful for potentially eliminating classes of biasing variables, whereas others might focus on a single bias.

Shadish et al. (2002) discussed an approach to building a case for causality that involves first establishing existence of a relationship between variables and then ruling out plausible alternatives. Such a programmatic approach seems most prudent for dealing with biases. First, one should establish that variables of interest are related, and this might be done most efficiently with a monomethod study, perhaps with self-reports if that is a reasonable means of assessing the variables of interest. Second, one should do a series of studies and analyses to control and test for plausible biases that might have distorted the observed relationship. This is likely to be a complex and difficult procedure to complete, particularly because in many areas, most developed and validated measures are self-reports. Furthermore, conducting longitudinal and multimethod studies can be expensive and labor intensive. Finally, there might not be the publication payoff with a follow-up study that merely shows with multimethod data what earlier monomethod studies already established.

The urban legend that there is universally shared variance inherent in our methods is both an exaggeration and oversimplification of the true state of affairs. Common method variance as often conceptualized may be a legend, but biases are real and endemic to our research. Furthermore, this is not just a problem of survey and field research. Even laboratory experiments have problems with bias, such as experimenter expectancies and demand characteristics

(Rosenthal & Rosnow, 1969). More sophisticated thinking about method variance is needed than the often knee-jerk complaints of CMV or monomethod bias we hear from both authors and reviewers. The time has come to retire the term *common method variance* and its derivatives and replace it with a consideration of specific biases and plausible alternative explanations for observed phenomena, regardless of whether they are from self-reports or other methods. Ruling out such alternatives through a program of systematic tests using a variety of methods will help establish the validity of conclusions based on initial monomethod studies.

References

- Boswell, W. R., Boudreau, J. W., & Dunford, B. B. (2004). The outcomes and correlates of job search objectives: Searching to leave or searching for leverage? *Journal of Applied Psychology, 89*, 1083-1091.
- Brief, A. P., Burke, M. J., George, J. M., Robinson, B., & Webster, J. (1988). Should negative affectivity remain an unmeasured variable in the study of job stress? *Journal of Applied Psychology, 73*, 193-198.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.
- Chan, D. (2001). Method effects of positive affectivity, negative affectivity, and impression management in self-reports of work attitudes. *Human Performance, 14*, 77-96.
- Chen, P. Y., Dai, T., Spector, P. E., & Jex, S. M. (1997). Relationship between negative affectivity and positive affectivity: Effects of judged desirability of scale items and respondent's social desirability. *Journal of Personality Assessment, 69*, 183-198.
- Chen, P. Y., & Spector, P. E. (1991). Negative affectivity as the underlying cause of correlations between stressors and strains. *Journal of Applied Psychology, 76*, 398-407.
- Connolly, J. J., & Viswesvaran, C. (2000). The role of affectivity in job satisfaction: A meta-analysis. *Personality and Individual Differences, 29*, 265-281.
- Crampton, S. M., & Wagner, J. A., III. (1994). Percept-percept inflation in microorganizational research: An investigation of prevalence and effect. *Journal of Applied Psychology, 79*, 67-76.
- Doty, D. H., & Glick, W. H. (1988, April). *Method variance in I/O research major effect or mythical beast?* Paper presented at the Society for Industrial and Organizational Psychology Conference, Dallas, TX.
- Doty, D. H., & Glick, W. H. (1998). Common methods bias: Does common methods variance really bias results? *Organizational Research Methods, 1*, 374-406.
- Frese, M. (1985). Stress at work and psychosomatic complaints: A causal interpretation. *Journal of Applied Psychology, 70*, 314-328.
- Frese, M., & Zapf, D. (1988). Methodological issues in the study of work stress: Objective vs. subjective measurement of stress at work and the question of longitudinal studies. In C. L. Cooper & R. Payne (Eds.), *Causes, coping, and consequences of stress at work* (pp. 375-411). Chichester, UK: Wiley.
- Glick, W. H., Jenkins, G. D., Jr., & Gupta, N. (1986). Method versus substance: How strong are underlying relationships between job characteristics and attitudinal outcomes? *Academy of Management Journal, 29*, 441-464.
- Lance, C. E., Baxter, D., & Mahan, R. P. (2005). Evaluation of alternative perspectives on source effects in multi-source performance measures. In W. Bennett Jr., C. E. Lance, & D. J. Woehr (Eds.), *Performance measurement: Current perspectives and future challenges*. Mahwah, NJ: Lawrence Erlbaum.
- Lance, C. E., Newbolt, W. H., Gatewood, R. D., Foster, M. R., French, N. R., & Smith, D. E. (2000). Assessment center exercise factors represent cross-situational specificity, not method bias. *Human Performance, 13*, 323-353.
- Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology, 86*, 114-121.
- Moorman, R. H., & Podsakoff, P. M. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behavior research. *Journal of Occupational and Organizational Psychology, 65*, 131-149.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660-679.

- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879-903.
- Podsakoff, P. M., & Todor, W. D. (1985). Relationships between leader reward and punishment behavior and group processes and productivity. *Journal of Management, 11*, 55-73.
- Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin, 63*, 129-156.
- Rosenthal, R., & Rosnow, R. L. (Eds.). (1969). *Artifact in behavioral research*. New York: Academic Press.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Thousand Oaks, CA: Sage.
- Spector, P. E. (1987). Method variance as an artifact in self-reported affect and perceptions at work: Myth or significant problem? *Journal of Applied Psychology, 72*, 438-443.
- Spector, P. E. (1994). Using self-report questionnaires in OB research: A comment on the use of a controversial method. *Journal of Organizational Behavior, 15*, 385-392.
- Spector, P. E., & Brannick, M. T. (1995). The nature and effects of method variance in organizational research. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 10, pp. 249-274). New York: John Wiley.
- Spector, P. E., Fox, S., & Van Katwyk, P. T. (1999). The role of negative affectivity in employee reactions to jobs: Nuisance effect or substantive effect? *Journal of Occupational and Organizational Psychology, 72*, 205-218.
- Spector, P. E., Jex, S. M., & Chen, P. Y. (1995). Personality traits as predictors of objective job characteristics. *Journal of Organizational Behavior, 16*, 59-65.
- Spector, P. E., Zapf, D., Chen, P. Y., & Frese, M. (2000). Why negative affectivity should not be controlled in job stress research: Don't throw out the baby with the bath water. *Journal of Organizational Behavior, 21*, 79-95.
- Watson, D., Pennebaker, J. W., & Folger, R. (1987). Beyond negative affectivity: Measuring stress and satisfaction in the workplace. In J. M. Ivancevich & D. C. Ganster (Eds.), *Job stress: From theory to suggestion* (pp. 141-157). New York: Haworth Press.
- Williams, L. J., & Anderson, S. E. (1994). An alternative approach to method effects using latent-variable models: Applications in organizational behavior research. *Journal of Applied Psychology, 79*, 323-331.
- Williams, L. J., & Brown, B. K. (1994). Method variance in organizational behavior and human resources research: Effects on correlations, path coefficients, and hypothesis testing. *Organizational Behavior and Human Decision Processes, 57*, 185-206.

Paul E. Spector is a professor of industrial/organizational (I/O) psychology and the I/O doctoral program director at the University of South Florida. His work has appeared in many journals, including the *Academy of Management Journal*, *Journal of Applied Psychology*, *Journal of Management*, *Journal of Organizational Behavior*, *Journal of Occupational and Organizational Psychology*, *Journal of Vocational Behavior*, *Organizational Behavior and Human Decision Processes*, *Personnel Psychology*, and *Psychological Bulletin*. At present, he is the Point/Counterpoint editor for the *Journal of Organizational Behavior* and is on the editorial boards of the *Journal of Occupational and Organizational Psychology*, *Organizational Research Methods*, and *Personnel Psychology*. In 1991, the Institute for Scientific Information listed him as one of the 50 highest impact contemporary researchers (out of more than 102,000) in psychology worldwide.