7-2010

What is Method Variance and How Can We Cope With It? A Panel Discussion

Michael T. Brannick
University of South Florida

David CHAN
Singapore Management University, davidchan@smu.edu.sg

James M. Conway
Central Connecticut State University

Charles E. Lance
University of Georgia

Paul E. Spector
University of South Florida,

DOI: https://doi.org/10.1177/1094428109360993

Published in ORGANIZATIONAL RESEARCH METHODS, 2010 July, 13 (3), 407-420
https://doi.org/10.1177/1094428109360993

What Is Method Variance and How Can We Cope With It? A Panel Discussion

Michael T. Brannick[1], David Chan[2], James M. Conway[3], Charles E. Lance[4], and Paul E. Spector[1]

Abstract

A panel of experts describes the nature of, and remedies for, method variance. In an attempt to help the reader understand the nature of method variance, the authors describe their experiences with method variance both on the giving and the receiving ends of the editorial review process, as well as their interpretation of other reviewers' comments. They then describe methods of data analysis and research design, which have been used for detecting and eliminating the effects of method variance. Most methods have some utility, but none prevent the researcher from making faulty inferences. The authors conclude with suggestions for resolving disputes about method variance.

Keywords

Method variance, Self-report, Construct validity, Survey research, Study design

Method variance is a problem both familiar and vexing to organizational researchers. What is method variance? Why is it a problem? What can we do about it? In the following discussion, the authors address eight such questions. Some of the eight questions are phrased in terms of the scholarly review process in refereed publications, as this is often where the problem gets discussed. However, the theoretical and practical issues are every bit as worthy of public scrutiny as they are of back channel criticism and argument. The following article provides the reader with expert opinions from a panel of four researchers who have written about the issues: David Chan, James M. Conway, Charles E. Lance, and Paul E. Spector.

The questions and discussion can be conceptualized in two broad areas: first, what is method variance? Panelists responded to questions about criticisms they have received from reviewers, criticisms that they have given as reviewers, and describe they think reviewers actually mean when

[1] Psychology Department, University of South Florida, Tampa, FL, USA
[2] School of Social Sciences, Singapore Management University, Singapore
[3] Psychology Department, Central Connecticut State University, New Britain, CT, USA
[4] Department of Psychology, University of Georgia, Athens, USA

Corresponding Author:
Michael T. Brannick, Psychology Department, PCD 4118G, University of South Florida, Tampa, FL 33620, USA
Email: mbrannic@luna.cas.usf.edu

they mention method variance. Second, panelists respond to questions regarding what to do about method variance. Questions under this heading include how to respond to reviewers' criticisms, two questions about analyzing data to detect and/or control method variance, and how to design studies in light of concerns about method variance. The concluding question asks panelists what they think should be done next to resolve disputes regarding method variance.

Questions and Responses

1. If you have received criticism from reviewers that your work suffered from method variance, what was the circumstance?

Lance: My dissertation data were all self-report survey data and among the other criticisms I faced from reviewers was that of common method (i.e., self-report) bias. I think that the issue of percept–percept bias (Crampton & Wagner, 1994) was at least as salient then (the mid-1980s) as it is today, partly as a result of Campbell's (1982) comments as outgoing editor of the *Journal of Applied Psychology* (*JAP*):

> . . . with perhaps one or two exceptions there has been very little opportunity to exercise any professional biases [as editor of *JAP*]. One possible exception pertains to the use of a self-report questionnaire to measure *all* the variables in a study. If there is no evident construct validity for the questionnaire measure or no variables that are measured independently of the questionnaire, I am biased against the study and believe that it contributes very little. (p. 692)

Because I used some fairly novel (at least at the time) analytic approaches I was finally able to emphasize these aspects of the study and publish it in a methodological journal (*Multivariate Behavioral Research*), but I could not get it into one of our substantive journals.

Conway: One instance in particular comes to mind. I had analyzed data from a number of multitrait–multirater studies of job performance and summarized the results; I found a relatively high correlation between task performance and contextual performance. One reviewer commented that this approach uses the same raters for each performance dimension, which tends to inflate relationships, and recommended using different raters for the different dimensions as an ideal strategy (the paper was published anyway).

Chan: As an author and editor, I can testify that the most common reviewer criticisms with regard to method variance and common method variance are alleged problems associated with the use of self-reports as the data collection method. The criticisms reflect the widespread belief among reviewers that self-report data reflect method variance rather than the intended true construct variance and they therefore have little construct validity. The criticisms also assume that the presence of method variance implies that self-report data are unable to provide accurate parameter estimates of inter-construct relationships. The consequence is that many reviewers will automatically dismiss manuscripts when self-report data were used in the study as the basis for substantive inferences.

2. If you have criticized someone's work for method variance in your role as reviewer, what was the circumstance?

Spector: I do not recall ever raising this criticism. As I have written elsewhere (Spector, 1977, 2006), method variance is an urban legend. The idea of method variance is that the method itself serves as a spurious cause of relationships among our variables. I've seen little if any credible evidence for this idea, and considerable evidence to the contrary, which is documented in my 1977 and

2006 papers. This is not to say that there aren't biases and spurious causes of observed correlations. It is just that they are not tied to methods themselves, but to the combination of methods and constructs (Spector & Brannick, 1995). If I am concerned about a specific design flaw or limitation, I will note that. For example, I might say something about not being able to draw causal conclusions from cross-sectional designs, or I might raise concern about item overlap in two scales that are supposed to assess different constructs. I might refer to particular potential biases, such as social desirability with sensitive topics.

Conway: I really have not had many experiences in which I pointed to the possible existence of method variance as a key issue with a manuscript (e.g., all variables measured by self-report, which could inflate relationships). Maybe this is because of the kinds of topics I'm asked to review. What I do see and comment on is the way people address method variance. I have reviewed quite a few studies using multitrait–multimethod (MTMM) designs and, while some are excellent, fairly often there are some basic flaws in execution (e.g., not using appropriate analysis, or not even reporting an MTMM analysis).

Lance: Most of the papers that I review these days or for which I serve as action Editor at *ORM* are more methodologically oriented, so method variance is not often an issue. However, I did not long ago review a paper in which the authors measured a relatively small number (6 or so) of macro-level organizational attributes using single- or few-item self-report measures. I raised the issue of common method variance in my review because even with small numbers of scale items, the variables' intercorrelations were in the .60s to .80s, which also led me to question the discriminant validity of the measures. For a combination of reasons (including theoretical ones), the paper was eventually rejected.

In another, similar case all study variables were measured by self-report except for one which was one of the study's main variables (a work environment characteristic) that was also measured by coworker report. All variables were conceptually similar and were (apparently) the only variables that were measured by the researchers. Not surprisingly, the self-report variables were all positively intercorrelated (mean $r = .44$) and the self-report predictors were correlated, on the average, twice as high with the self-reported work environment characteristic (mean $r = .33$) as compared to the coworker report measure (mean $r = .16$). Combined, the data collection procedures, the conceptual similarity of the variables, and the correlational evidence led me to question seriously whether percept–percept bias could serve as a possible alternative explanation for at least some of their findings.

Chan: When I raised the concern of method variance in my role as reviewer or editor, the circumstance was often about (a) a specific instance of systematic measurement error or (b) an error of substantive inference due to using an inadequate study design to test the hypothesis or support a substantive conclusion. An example of (a) is the presence of a predictor-related criterion bias such as using interview scores to predict supervisory ratings of job performance when the supervisor who produced both the interview ratings and job performance ratings may partly be influenced by rater–ratee similarity or ratee likeability that are conceptually unrelated to the constructs intended to be assessed by the predictor and criterion measures. The same bias could occur when supervisors serve as assessors in assessment centers and the assessor ratings are used to predict supervisory ratings of job performance. The rater–ratee similarity and the ratee likeability are systematic errors associated with the method of measurement. These errors represent method variance and contribute to the observed variance in each measure thereby lowering each measure's construct validity. These errors also represent common method variance and contribute to the observed covariance thereby artificially increasing the criterion-related validity. An example of (b) is the failure to make the method–content distinction when comparing scores between measures. This happens very often in studies of applicant reactions where the researcher directly compares test reaction scores between different types of selection measures such as structured interviews and paper-and-pencil cognitive ability tests. The two measures differ in both the method of testing and the test content. Because the test reaction items refer to the test without specifying whether it is the method of testing or the test content that is the referent, test method variance and test content variance are confounded in the test reaction scores and direct comparisons of these scores and subgroup differences in these scores become problematic since method variance and content variance are not isolated.

3. When a reviewer complains about method variance, what is the essence of the complaint? What do you think they really mean?

Conway: I think our general notion of the problem with method variance is that (a) there is systematic error variance in a set of measurements, and (b) that systematic error variance may be shared with measurements on another variable, (c) resulting in an inflated estimate of the relationship. This systematic distortion of a relationship could be referred to as method *bias*. I think that's basically it. This might be salient, for example, in relationships between self-reported job characteristics and outcomes like motivation or satisfaction, or between different aspects of job performance rated by the same source (e.g., supervisor).

This isn't a response to the question I was asked, but I'm going to throw in a couple of additional comments anyway. One comment is that we (reviewers, authors, everyone) need to become more sophisticated in the way we conceptualize method variance. This should include thinking carefully about what variables may cause method effects in a given case. Larry Williams has developed a technique ("measured method effects") for partialling out method effects using structural equation models, but this can only work if we think carefully in advance about what types of method effects are likely to be present. Another layer is considering whether a given effect is really a method effect or a substantive effect. Paul Spector's work on negative affect in stress research is an excellent example. Negative affect has been considered a "method construct," adding irrelevant variance to stress measures, but it may be a substantively important variable; if so, controlling for it will remove valid variance from stress measures. Another excellent example is Chuck Lance's work on method factors in assessment center MTMM studies—he has argued that what show up as "method" factors really represent substantively important variance in assessment center ratings.

Another nonsanctioned comment is that method variance can attenuate relationships as well as inflate them. This isn't the way we normally think about things, but I think it's potentially just as problematic. An example I've used in the past involves direct report ratings of a supervisor's behavior and an objective index of financial success. These two variables are unlikely to share common method variance. However, each is likely to be affected by some method variables, adding irrelevant variance. The unshared irrelevant variance will attenuate the relationship unless somehow taken into account. I wonder to what extent our literature is affected by underestimated relationships.

A final comment is that we really don't have a good handle on how big a problem method variance, and inflation due to shared method variance, is. There has been a lot of MTMM work, and some of it suggests that there is generally relatively small inflation. But problems with how we analyze MTMM data (e.g., assumptions required for different confirmatory factor analysis [CFA] models) make these results open to question.

Lance: One of two things, either that the measure's construct validity has been compromised by method variance or that measures' covariances are distorted by *common* method variance (CMV). The first idea is illustrated from a simple extension of classical test theory:

$$X_{ij} = T_{\mathrm{i}} + M_j + E_{ij}, \tag{1}$$

where some observed measure X_{ij} reflects the influences of the corresponding ith true score (T_i) and the jth measurement method (M_j), and nonsystematic measurement error (E_{ij}). Assuming that all terms are expressed in deviation form, and that $E(T_i,M_j) = E(T_i,E_{ij}) = E(M_j,E_{ij}) = 0$, where $E()$ is the expected value operator, then

$$\sigma^2_{X_{ij}} = \sigma^2_{T_i} + \sigma^2_{M_j} + \sigma^2_{E_{ij}}. \tag{2}$$

Thus, to the extent that method effects are prominent, a greater proportion of the variance in X_{ij} is attributable to the method of measurement $\left(\sigma^2_M\right)$ relative to the intended construct $\left(\sigma^2_T\right)$ and construct validity is compromised. Thus, I think this is one concern that reviewers have—that if method effects are prominent then σ^2_X reflects proportionally more the influence of σ^2_M relative to σ^2_T.

The second idea, covariance distortion due to CMV, can be illustrated by first rewriting Equation (1) so that the relative contributions of T_i and M_j are acknowledged more explicitly:

$$X_{ij} = \lambda_{T_{ij}} T_i + \lambda_{M_{ij}} M_j + E_{ij}, \tag{3}$$

where the λs represent relative (regression) weights. Then, under the same assumptions as before, the covariance between measures of two different constructs measured by the same method (a heterotrait-monomethod covariance in MTMM nomenclature) can be written as:

$$\sigma_{X_{ij}X_{i'j}} = \lambda_{T_{ij}}\lambda_{T_{i'j}}\phi_{T_iT_{i'}} + \lambda_{M_{ij}}\lambda_{M_{i'j}}, \tag{4}$$

where $\phi_{T_iT_{i'}}$ represents the covariance between constructs T_i and $T_{i'}$ and $\lambda_{M_{ij}}\lambda_{M_{i'j}}$ represents a source of covariance between the *measures* of T_i and $T_{i'}$ that results when both variables are measured using method M. Thus, if CMV is prevalent, the observed covariance between X_{ij} and $X_{i'j}$ will be inflated by a factor equal to $\lambda_{M_{ij}}\lambda_{M_{i'j}}$. Note however, that the observed score covariance $\left(\sigma_{X_{ij}X_{i'j}}\right)$ is the true score covariance $\left(\phi_{T_iT_{i'}}\right)$ *attenuated* by $\lambda_{T_{ij}}\lambda_{T_{i'j}}$. As such, observed monomethod covariances are simultaneously *inflated* by CMV and *attenuated* by measurement error, but I don't think this is widely recognized. Lance and Sloan (1993) showed how the estimated latent trait correlations in CFA of MTMM data are simultaneously corrected for disattenuation due to unreliability and spuriousness due to CMV:

$$\phi_{T_iT_{i'}} = \frac{r_{T_{ij}T_{i'j}} - \lambda_{M_{ij}}\lambda_{M_{i'j}}}{\lambda_{T_{ij}}\lambda_{T_{i'j}}}. \tag{5}$$

Using this approach in an MTMM study of life satisfaction ratings made using different scale formats, Lance and Sloan (1993) found, perhaps coincidentally, that the attenuating effects of measurement error almost exactly offset the inflationary effects of CMV: the average observed monomethod correlation (mean $r = .30$) was almost identical to the average satisfaction trait factor correlation (mean $\phi = .29$). So, reviewers' second concern is most likely of the inflationary effects of CMV but these can be offset by the attenuating effects of measurement error so that the observed covariances may not be so distorted from their true score counterparts after all. However, just how widespread these counteracting influences are is unknown.

There is even a third possible concern that probably is not often voiced by reviewers and that is the influence of *related* measurement methods. Note that even if *different* measurement methods are used, covariance distortion due method effects can still incur:

$$\sigma_{X_{ij}X_{i'j'}} = \lambda_{T_{ij}}\lambda_{T_{i'j'}}\phi_{T_iT_{i'}} + \lambda_{M_{ij}}\lambda_{M_{i'j'}}\phi_{M_jM_{j'}}, \tag{6}$$

where $\phi_{M_jM_{j'}}$ refers to the covariance between the jth and the j'th measurement methods, and all other terms are defined as in Equation 4 (the covariance in Equation 6 is a heterotrait-heteromethod covariance in MTMM nomenclature). Thus, even when different measurement methods are used, if the methods are (highly) related (e.g., two parents' reports of children's behavior) and have significant effects on their measures, inflationary effects can still incur. In the event that the measurement methods are (essentially) unrelated (i.e., $\phi_{M_jM_{j'}}$ is near 0, e.g., paper-and-pencil test vs. peer ratings), then the construct validity of the measures could be compromised (see Equation 2) without the inflationary CMV. If, however, the covariance between the methods is *negative* (e.g., effects of positive vs. negative item wording) the effect of $\lambda_{M_{ij}}\lambda_{M_{i'j'}}\phi_{M_jM_{j'}}$ would be to *attenuate* the correlations between the observed measures. My guess is that the inflationary effects of CMV illustrated in Equation 4 are what reviewers are most concerned about. I do not think that the more subtle effects associated with Equation 6 are as widely acknowledged or appreciated.

Spector: I sometimes wonder if the reviewer has really thought through the issues with the study or if it is a knee-jerk reaction to a self-report study. The reviewer has seen this criticism so often that he or she just raises it automatically. I would like to think that this is a shorthand concern that the design is cross-sectional and can't lead to confident causal conclusions, or that there are likely biases shared by some of the measures in the study. It would be more helpful if reviewers were more precise in their criticisms.

4. Suppose you have completed a study (perhaps an organizational survey) and submitted it for review. You receive a complaint from a reviewer that your interpretation of the relations between variables could be spurious and due to method variance. What is the best response to such a criticism?

Spector: An argument can easily be made that it isn't "method variance." Just point to correlations in the study that are nonsignificant and near zero, or if you don't have any, look to the literature to show that your variables sometimes fail to relate to other self-report variables. The Lindell and Whitney (2001) approach can be useful for this purpose. Show that your variables are not related to social desirability, or to negative affectivity. What cannot be done so easily is demonstrating that a particular relationship is not due to third variables that have little to do with method. You can rule out some obvious ones, like social desirability, negative affectivity, or mood, assuming you built them into the study, or you can look to the literature to show that your variables are not related to these. But you can't think of everything that might potentially affect your results. Also, keep in mind that you can rule out something as a spurious cause of something, but it is not easy to show that something did cause spurious relationships. Spuriousness is a causal conclusion, that is, that the "cause" of the relationship between X and Y was due to Z because Z "causes" both X and Y. Just as you can't demonstrate causality by just showing X and Y are related, you can't demonstrate causality by just showing Z relates to both X and Y, or that the correlation between X and Y reduces substantially or to nonsignificance when Z is added to the analysis.

Conway: Well, the best response will depend on the situation, but it's best to address this type of issue up front, in the original submission. This requires a thoughtful approach to method variance which should improve design and analysis. But, if it does come up in a review, there are a number of potentially appropriate responses. One type might be conceptual—an author might argue that they have considered plausible method variables and have theoretical reason to believe that shared variance, inflating a relationship, is unlikely. But given our current state of knowledge of method variables, a purely theoretical argument probably won't fly. A second approach might be to argue that the research design precludes shared method effects, and a third approach is to present some type of data showing that shared method variance is not a plausible explanation for a relationship.

Lance: That depends on a number of things, perhaps the most important of which are (a) whether one believes that one's data *do* indeed suffer from CMV and (b) the kinds of evidence available to refute the reviewer's claim. Assuming that one does *not* believe that CMV is a serious threat, then one might (a) argue that some design considerations, for example, temporal separation of data collection, clearly labeling separate sections of a survey and providing definitions for separate variables, etc., should mitigate against CMV; (b) show that correlations between variables measured using the same method are not significantly higher than correlations between variables measured using different methods; (c) if MTMM analyses are possible, conduct CFA of the matrix and demonstrate low method factor loadings and/or correlations; (d) point to near-zero correlations between study variables and method-related variables (e.g., positive/negative affect, social desirability, etc.).

5. Is there a particular method of data analysis that you would recommend as a means of demonstrating that method variance is not of great concern for a given data set?

Chan: I don't think there is one best data analysis technique. The appropriateness of the technique is dependent on the hypothesis, study design, nature of the constructs, the type of measures used in the study, properties of the measurement methods, and specific substantive inference that we want to make from the results. If there is a good theory of the inter-construct relationships and independent measures (distinct from the focal measures that are allegedly suffering from method variance) of the specific method factor are available, then the latent variable modeling approach suggested by Williams, Cote, and Buckley (1989) provides a useful technique for decomposing the variance of each item on the focal measures into true construct variance, method variance due to one or more contaminating method factors, and random measurement error variance. This method is powerful and flexible as it allows us to specify and test various alternative models of method variance as well as estimate the method effects in each focal measure and the extent to which the method effects impact the parameter estimates of the substantive relationships between the constructs represented by the focal measures. Several studies applied this technique and showed that method variance due to factors such as ratee likeability, rater–ratee similarity, positive and negative affectivity, and impression management could occur in data sets but the impact of these method effects on the estimation of the substantive relationships involving the focal constructs may nevertheless be small or even trivial (Chan, 2001, 2008; Schmitt, Pulakos, Nason, & Whitney, 1996).

Conway: A variety of data analysis techniques can be useful, and it depends on the situation. For example, an MTMM analysis is an option if all constructs can be measured with multiple methods, but this isn't always possible.

Partialling approaches. One approach I do not recommend is partialling a first factor from a set of self-report measures; you don't really know what you're getting rid of.

Assessing relationships with method variables. Another approach, more defensible, is to assess whether substantive variables are related to plausible method variables such as social desirability or acquiescence. This can rule out particular method effects but only those which the researcher measures. One virtue of this approach is that it forces the researcher to conceptualize potential method problems in advance.

Measured method effects in a structural equation model. Larry Williams's approach (which I mentioned earlier) is a more sophisticated approach to measuring and including method variables in the design. The idea is to test a structural equation model in which hypothesized method variables act as causes of substantive variables, and their variance is removed from substantive variables. The result is a reduction in bias of relationship estimates. I would note that this approach is effective against both inflation due to shared method effects, and attenuation due to unshared method effects.

MTMM analysis. This is the area in which I've published the most, so I appreciate some of the benefits and problems. MTMM can be a nice way to both assess the extent of method effects and to remove them from estimation of relationships (this approach can be helpful with both inflation and attenuation). One important problem is that there is no perfect way to assess MTMM data. CFA is often used, but the model many see as closest to ideal (correlated trait factors and correlated method factors) often is difficult to estimate unless you have a very large sample and a large number of degrees of freedom. The correlated uniqueness model is much easier to estimate but requires the assumption of uncorrelated methods, which has been demonstrated to bias the relationships among traits. This approach requires careful planning, and a good place to start exploring it is to read the classic paper by Campbell and Fiske (1959).

Lance: Probably CFA of MTMM data is the most straightforward and powerful approach. One can quantify the strength of the (unmeasured) method effects and even conduct tests of hypothesized structural relationships among the substantive latent variables corrected for attenuation due to unreliability and purged of method effects (though I've seen very few examples of this). The approach of Williams and Anderson (1994) using substantive method-related latent variables and the partial correlation approach of Lindell and Whitney (2001) are also intriguing alternatives, but these also seem not to have been widely adopted or applied. I think some simulation work comparing these approaches would be valuable in highlighting their relative advantages, disadvantages, and effectiveness.

Spector: Again, you can easily show that method variance isn't a problem by showing that not everything in your study was correlated. But this does not rule out the possibility that certain pairs of variables were biased by a common cause and that relationships among them are spurious.

6. If you are worried about method variance before you collect data, what sort of design elements might you include to reduce or eliminate the threat of method variance?

Spector: I wouldn't be worried about method variance because I see it as an urban legend—something that people have heard so often they erroneously believe it exists (thanks to Bob Vandenberg for coming up with the term). Rather, I might be worried about shared biases among variables. I would use a number of strategies:

a. Include measures of suspected biases, like social desirability or mood.
b. Use multiple sources of data, although sources other than subjects of the study can share the same biases.
c. Use experimental or quasiexperimental designs where one or more variables were manipulated.
d. Include variables that are as factual as possible, so bias is unlikely. For example, working hours leaves little room for subjective interpretation and has very high convergence among sources (Spector, Dwyer, & Jex, 1988).
e. Use a longitudinal design where data are collected before and after some meaningful event of interest, for example, before and after the announcement of a downsizing. Keep in mind that two arbitrary points in time are not very helpful, although they can help rule out transient occasion factors such as daily mood.

Chan: One design approach, if the context permits, is to experimentally manipulate the method of data collection (i.e., test method) while keeping test content constant and vice versa. When test content differs but test method is kept constant, then the difference in scores between the two tests cannot be attributed to method variance although method variance may still occur in each test. Conversely, when test method differs but test content is kept constant, then the difference in scores between the two tests can be attributed to method variance. In a study of the importance of the method–content distinction in understanding racial subgroup differences in situational judgment test scores (Chan & Schmitt, 1997), we adopted a race-by-method between-subjects quasiexperimental design that manipulated the method of testing (paper-and-pencil vs. video) while keeping test content constant in order to isolate method variance and true construct variance. The factorial design allowed us to demonstrate a race by method interaction effect. We also included an independent measure of the method factor (i.e., reading comprehension) and showed that the interaction effect disappeared when the hypothesized method variable was statistically controlled thereby supporting the nature of the method variance factor. Although we did not, we could also have manipulated test content. I think the combined use of experimental design to manipulate test method (and/or test content) and an independent measure of the hypothesized method factor provides a rigorous method to not only eliminate or reduce the threat of method variance but also one that helps us understand the nature of the particular method variance in a given study and test the specific theory of method variance and estimate the magnitude of the method effects and their implications on the substantive inferences.

Conway: Another approach would be to develop measures less susceptible to method effects. This is one of the things that Campbell and Fiske (1959) had in mind when they proposed the MTMM matrix—it should be a way to help us improve our measures rather than just partialling out method effects. One example of work in this direction is Paul Spector's Factual Autonomy Scale, and another is Larry James' work on conditional reasoning to measure personality—but overall, we really haven't moved very far in that direction.

Lance: It is instructive to see what kinds of alternative methods were being considered at the time of the original paper by Campbell and Fiske (1959), for example, various apparatuses (e.g., "Obstruction Box" vs. "Activity Wheel" measurements of "drive" in animals), peer ratings versus paper-and-pencil tests of personality traits, sociometric ratings versus observation by self and others of popularity and expansiveness, "free behavior" role play and projective test measurement of interaction process variables, various test formats, and raters (e.g., staff, teammate, and self-ratings clinical symptoms). In the last few decades, the MTMM methodology has been applied in I/O psychology mainly to study a wide variety of "traits" by methods operationalized as different (a) raters/rater groups (e.g., 360° performance ratings), (b) test forms (e.g., CPI vs. 16PF measures of the Big 5), (c) exercises in assessment centers (e.g., role-play vs. in-basket), (d) scale formats (e.g., forced choice vs. Likert type), and (e) occasions of measurement (see Lance, Baranik, Lau, & Scharlau, 2009). Now, returning to the question, the effects of some of these alleged "methods" probably *cannot* be reduced as they reflect robust substantive effects on measured variables, and not mere nuisance method effects. For example, we have argued and documented that rater (source) effects represent different but perhaps equally valid perspectives on ratee performance, and not performance-irrelevant rating error (Lance, Hoffman, Baranik, & Gentry, 2008). As two other examples, exercise effects in assessment centers have been shown to represent true cross-situational specificity in performance, not method bias effects (Lance, 2008), and occasion factors have been shown to reflect state aspects of the traits being measured (Schermelleh-Engel, Keith, Moosbrugger, & Hodapp, 2004). Thus, one of the questions that we urge researchers to consider is whether some alleged "method" is merely a procedure for assigning numbers to variables in order to operationalize constructs or whether some substantively meaningful influences are also involved.

7. When two variables are measured through a self-report questionnaire (e.g., organizational commitment and job satisfaction), some may object that the interpretation of the observed correlation between the two is spurious because it really indicates the influence of some (typically unspecified) third variable. Three main approaches to rebutting the argument that observed relations are spurious are
   (a) pairing a self-report variable with a variable that is not-self-reported;
   (b) removing the effect of specific, hypothesized alternative explanatory variable (e.g., partial negative affectivity from commitment and job satisfaction); and
   (c) general partialling methods (e.g., partial the first principal component, subtract the lowest observed correlation among the set of variables measured in the survey).
      i. Comment on any of these methods.
      ii. Are there other methods that you think are superior to these? Please describe.

Chan: There are situations in which it is worse to use non-self-report measures than self-report measures to measure the same intended constructs. For example, the use of self-report measure is not only justifiable but also probably necessary when assessing constructs that are self-referential perceptions such as job satisfaction, perceived organizational support, and perceived fairness. To find out about the perception of an individual, it is often best to ask the individual about his or her perception rather than infer it indirectly from what others observe about the individual's behaviors. For self-referential perception constructs, using other reports is problematic for at least three reasons. First, the individual's perceptions may not translate into observable behaviors. Second, even if perceptions were translated into behaviors, others may not have the opportunity to observe these relevant behaviors. Third, valid measurement by other reports requires the reporter to accurately infer the individual's specific perception and the specific value on that perception from the observation of the individual's behavior. In short, when assessing self-referential perception constructs, non-self-report measures are often inferior in validity when compared to self-report measures.

It is also not true that we can always be more confident of the validity of a self-report measure if the scores converge with the scores on the corresponding non-self-report measures. Both artificial inflation of correlation due to predictor-related criterion bias and artificial deflation of correlation due to suppressor effect may also occur when a self-report measure is correlated with the corresponding non-self-report measure. Consider the correlation between a self-report measure of conscientiousness and a supervisor-report measure of conscientiousness. If both measures were affected by impression management of the rated individuals, then artificial inflation of correlation would occur. On the other hand, if impression management of the rated individual affected the self-report measure but not the supervisor-report measure, then a suppressor effect and hence artificial deflation of correlation would occur. This implies that the observed correlation between a self-report measure and the corresponding non-self-report measure is not necessarily a good indication of the validity of the self-report measure since the correlation may be artificially inflated or deflated.

Spector: I noticed that you phrased this question, not as a way of addressing method variance per se, but as a means of dealing with a potential third variable. I think this is an important distinction, as the idea of method variance is that method itself is the third variable, and that it affects all variables assessed with that method.

Approach 1 is a reasonable approach although it has its limitations. Keep in mind that the typical approach of using supervisor, coworker, or observer ratings is, if anything, potentially more biased and less accurate than subject self-reports (Dalal, 2005; Frese & Zapf, 1994). If you find convergence of results between self-reports and non-self-reports, you will have confidence that it wasn't biases within the self-reports that accounted for results. If, however, results differ, you won't know if it is due to biases within the self-report, or inaccuracy in the alternative source of data.

Approach 2 is another reasonable thing to try in an attempt to rule out suspected biases. Again, failing to rule out the potential bias is not evidence that it is a bias. Just because the partial correlation (or regression coefficient) reduces to nonsignificance doesn't indicate that the NA is the "cause" of the correlation between commitment and job satisfaction.

Approach 3 might be helpful if method is the third variable and has an effect on all variables assessed with it. Since that is not the case, or at least there is little if any credible evidence that it is, the approach is not worth considering, other than perhaps as a way of getting past a reviewer.

8. What do you think is the most constructive thing that could be done now to resolve discussions and disputes about method variance?

Lance: Recognize that (a) variation on methods of measurement represent but one facet in a potentially multidimensional data array and (b) simply because one measurement facet is not the focal measurement facet, does not necessarily make it a measurement method facet. As I mentioned earlier, many different variations have been proffered as alternative measurement methods including different rater sources (e.g., peers vs. supervisors vs. subordinates, etc.), different response formats (e.g., semantic differential, Likert-type items, etc.), different tests of the same construct (e.g., the Hogan vs. the 16PF vs. the CPI as Big 5 measures), and so on. In a recent review of MTMM studies (Lance et al., 2009), we argued that there seems to be a widespread default assumption that the measurement facet that is *not* the one of interest (e.g., different job performance dimensions, Big 5 personality dimensions, assessment center dimensions, i.e., the "Traits" of interest), must therefore constitute the measurement method facet in a MTMM sense. We suggested that researchers should view their data structures in the context of a prototype multidimensional measurement system consisting of (a) persons (or groups or collectivities), (b) focal constructs, (c) occasions, (d) situations, (e) observers or recorders, and (f) response modalities (see Lance et al., p. 354) in order to locate their data structures more meaningfully within this larger structure and to help clarify which facets might properly be interpreted as a measurement method facet and which other facets might properly be construed as representing something (much) more than mere method.

Spector: Retire the term "method variance" from our vocabulary and instead talk about the real issues in a more precise way. I prefer to talk about biases that affect particular sets of variables, think about the limits of what people can and cannot self-report accurately, and consider what can and cannot be concluded from data collected using particular designs and methods. It is perfectly reasonable at early stages of research to establish that two or more variables of interest are related, using self-reports. Once we establish the relationship, the next step is to figure out why they are related. This requires a series of studies using a variety of methods to rule in or rule out potential biases and confounds and to establish if observed relationships are causal. There are no easy answers, and ultimately all we can do is build our case with a series of studies.

The idea of method variance has served its purpose in sensitizing the field to issues of bias in assessment, and in not taking results at face value. But its time has passed. I see no useful purpose in applying one of the many method variance tests to show that it wasn't a problem in a particular study, other than to get past a reviewer and get a paper published. Podsakoff, MacKenzie, Lee, and Podsakoff (2003) have done a marvelous job of documenting a long list of such "reviewer silencing" tests. I don't underestimate the value of having some useful procedures to address reviewer criticism, but their use is scientifically counterproductive as it leads to a false sense of security that everything is fine with a study because there was no method variance. And I see no value in continuing to conduct research on methods to control for method variance. You cannot control for what doesn't exist. Let's get a better handle on the various things that bias measurement and serve as third variables that influence observed correlations. Our time would be much better spent designing systematic strategies to address our research questions in a more comprehensive and convincing way that explores all sorts of potential biases and third variables than the fictitious common method variance.

Chan: I think some of the problems of method variance are overstated or exaggerated. We need a change of mind-set in our approach to method variance, especially with regard to self-report data. For example, we should not take as default mode the position that self-report data are inherently full of serious problems of method variance that automatically lead to fallacious inferences and hence require a rejection decision on manuscripts. Instead, we should examine how the pros and cons of self-report data may apply to a given study, similar to how we would evaluate any other types of measurement method and data source.

Method variance is simply that part of the observed variance that is due to factors associated with the method of testing that are distinct from the focal construct that the test or measure was designed to assess. The presence of method variance per se does not necessarily imply a fatal flaw in the study. We need to evaluate the criticisms about method variance, and to do so would require us to be explicit about the intended and unintended constructs represented by the measures, as well as the substantive content of the items on the measure.

Conway: In my opinion, the most important thing to do is be more careful about conceptualizing method variance. We should do this as a field and as individual researchers. Research to give us a better understanding of what method variables cause method variance would help us do a much better job of controlling it, either statistically or by developing better measurement methods.

Summary

Brannick: The concluding comments represent my understanding of method variance in light the panelists' responses to the eight posed questions. What is method variance? One answer to such a question is that method variance is an umbrella or generic term for invalidity of measurement. Systematic sources of variance that are not those of interest to the researcher are good candidates for the label "method variance." Although method variance is often invoked in the context of surveys (self-reports), it is not limited to such reports, but is rather a potential concern for any measure, because validity of measurement is always important. Method variance matters because it represents an alternative explanation of substantive results. That is, method variance could lead the researcher into faulty inferences regarding substantive questions of interest.

The panelists suggested that careful thought be given to specific sources of invalidity when designing studies rather than assuming that any given procedure of measurement (e.g., self-report surveys) automatically produces a certain amount of method variance. The panelists also referenced a taxonomy of features of methods (Lance et al., 2009) that might prove useful in thinking about method variance, though the list is not exhaustive of threats to the validity of measurement. Despite its problems, self-report is sometimes the preferred method of measurement.

There are many statistical techniques that can be used to rebut reviewer criticisms of method variance. There are also many statistical approaches to detecting and controlling method variance. Although such techniques have merit under some circumstances, none of them provides fail-safe protection against faulty inferences. After all, validation is an ongoing enterprise that results in better and better understanding of the meaning of observed scores. The panelists cautioned that the smallest correlation in a correlation matrix of measures using the same method does not necessarily indicate the amount of invalidity in the correlation of substantive interest, nor does the general level of correlations in the matrix or the first factor (or principal component) from the matrix. Although latent variable approaches to method variance have promise, they also have problems, particularly when applied to MTMM data.

If method variance is considered before data collection, then the researcher can design the study in ways to minimize threats to validity based on measurement methods. Specific hypothesized threats such as negative affectivity can be included as measured variables in the study. Features of method can be manipulated experimentally (e.g., Chan & Schmitt, 1997). Clearly, not every possible threat to validity can be explicitly controlled in a single study.

In response to the question regarding what to do next, the panelists called for a new mind-set regarding method variance. Rather than considering method variance to be a plague, which, once contracted by a study, leads inevitably to death (read: rejection of publication), method variance should be regarded in a more refined way. Put another way, rather than considering method variance to be a general problem that afflicts a study, authors and reviewers should consider specific problems in measurement that affect the focal substantive inference. Thus, specific sources of irrelevant systematic variance that could produce a faulty inference should be considered and reduced or eliminated, preferably by design.

In summary, method variance remains a serious concern in the conduct of organizational research. However, the more complex way of thinking about method variance that is suggested by the panelists should prove helpful in producing better research and better understanding of human behavior in organizations.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Funding

The authors received no financial support for the research and/or authorship of this article.

References

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

Campbell, J. P. (1982). Editorial: Some remarks from the matrix. *Journal of Applied Psychology*, *67*, 691-700.

Chan, D. (2001). Method effects of positive affectivity, negative affectivity, and impression management in self-reports of work attitudes. *Human Performance*, *14*, 77-96.

Chan, D. (2008). So why ask me?—Are self-report data really that bad? In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Received doctrine, verity, and fable in the organizational and social sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, *82*, 143-159.

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.

Dalal, R. S. (2005). A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *Journal of Applied Psychology*, *90*, 1241-1255.

Frese, M., & Zapf, D. (1994). *Methodological issues in the study of work stress: Objective vs subjective measurement of work stress and the question of longitudinal studies.* Oxford, UK: John Wiley.

Lance, C. E. (2008). Why assessment centers do not work the way they are supposed to. *Industrial and Organizational Psychology*, *1*, 84-97.

Lance, C. E., Baranik, L. E., Lau, A. R., & Scharlau, E. A. (2009). If it ain't trait it must be method: (Mis)application of the multitrait-multimethod methodology in organizational research. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Received doctrine, verity, and fable in organizational and social research* (pp. 339-362). New York: Routledge.

Lance, C. E., Hoffman, B. J., Baranik, L. E., & Gentry, W. A. (2008). Rater source factors represent important subcomponents of the criterion construct space, not rater bias. *Human Resource Management Review*, *18*, 223-232.

Lance, C. E., & Sloan, C. E. (1993). Relationships between overall and life facet satisfaction: A multitrait-multimethod (MTMM) study. *Social Indicators Research*, *30*, 1-15.

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

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.

Schermelleh-Engel, K., Keith, N., Moosbrugger, H., & Hodapp, V. (2004). Decomposing person and occasion-specific effects: An extension of latent state-trait theory to hierarchical models. *Psychological Methods*, *9*, 198-219.

Schmitt, N., Pulakos, E. D., Nason, E., & Whitney, D. J. (1996). Likability and similarity as potential sources of predictor-related criterion bias in validation research. *Organizational Behavior and Human Decision Processes*, *68*, 272-286.

Spector, P. E. (1977). What to do with significant multivariate effects in multivariate analyses of variance. *Journal of Applied Psychology*, *62*, 158-163.

Spector, P. E. (2006). Method variance in organizational research: Truth or urban legend? *Organizational Research Methods*, *9*, 221-232.

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: 1995* (pp. 249-274). West Sussex, UK: John Wiley.

Spector, P. E., Dwyer, D. J., & Jex, S. M. (1988). Relation of job stressors to affective, health, and performance outcomes: A comparison of multiple data sources. *Journal of Applied Psychology*, *73*, 11-19.

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

Williams, L. J., Cote, J. A., & Buckley, M. R. (1989). Lack of method variance in self-reported affect and perceptions at work: Reality or artifact? *Journal of Applied Psychology*, *74*, 462-446.

Bios

Michael T. Brannick obtained his PhD from Bowling Green State University and is currently Professor and Interim Chair of the Psychology Department at the University of South Florida. He has authored over 50 articles on a variety of topics including research methods and performance measurement.

David Chan received his PhD from Michigan State University. He is currently Professor of Psychology and Vice Provost (Research & Graduate Studies) at Singapore Management University. He has published numerous journal articles, and has received several scholarly awards including the Distinguished Early Career Contributions Award, the William Owens Scholarly Achievement Award, and the Edwin Ghiselli Award for Innovative Research Design presented by SIOP. He has served on many editorial boards, including those for *Personnel Psychology, Human Performance,* and *Organizational Research Methods*.

James M. Conway received his PhD from the University of Connecticut, and is currently the Assistant Chair of the Psychology Department at Central Connecticut State University. He has published numerous articles regarding quantitative methods, particularly regarding the analysis of the Multitrait-Multimethod Matrix, and also regarding ratings of job performance and *interview performance.*

Charles E. Lance is a Professor of I/O psychology at The University of Georgia. He has published extensively on measurement methods; one indication of the quality of this work is the awards he has won from the Academy of Management for papers on research methods. He is currently Associate Editor of *Organizational Research Methods*, and on the editorial boards of *Personnel Psychology, Human Performance,* and *Group & Organization Management*.

Paul E. Spector is a Distinguished Professor of psychology and the Director of the Industrial/Organizational Psychology Doctoral Program at the University of South Florida. He has published more than 100 articles in peer reviewed journals, and has served on editorial boards for many of the major I/O journals including *Journal of Applied Psychology, Journal of Management, Organizational Research Methods,* and *Personnel Psychology*. He has published several articles and chapters on method variance and related issues.