Organizational Research

Methods

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

**Method Effects, Measurement Error, and Substantive Conclusions**

Charles E. Lance, Bryan Dawson, David Birkelbach and Brian J. Hoffman

*Organizational Research Methods* 2010 13: 435 originally published online 28 January 2010

DOI: 10.1177/1094428109352528

The online version of this article can be found at:

<http://orm.sagepub.com/content/13/3/435>

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/content/13/3/435.refs.html>

>> [Version of Record -](http://orm.sagepub.com/content/13/3/435.full.pdf) Jun 21, 2010

[OnlineFirst Version of Record -](http://orm.sagepub.com/content/early/2010/01/28/1094428109352528.full.pdf) Jan 28, 2010

What is This?

Downloaded from [orm.sagepub.com at](http://orm.sagepub.com/) UNIV OF GEORGIA LIBRARIES on October 9, 2013

Method Effects, Measurement Error, and Substantive Conclusions

Charles E. Lance,1 Bryan Dawson,1

David Birkelbach,1 and Brian J. Hoffman1

Organizational Research Methods

13(3) 435-455

ª The Author(s) 2010

Reprints and permission: sagepub.com/journalsPermissions.nav DOI: 10.1177/1094428109352528 <http://orm.sagepub.com>



Abstract

Common method variance is routinely viewed as a pervasive problem in organizational research, one that undermines good science and biases empirical conclusions. The authors review research that has used multitrait multimethod (MTMM) designs to estimate the magnitude of common method variance in organizational research. The results of this study show that method variance accounts for less variance (18%) than has been suggested by previous reviews. The authors also consider simultaneously the attenuating effect of measurement error with the inflationary effect of common method variance on observed relationships. Results indicate that although common method variance does have an inflationary effect on observed relationships, this effect is almost completely offset by the attenuating effect of measurement error.

Keywords

measurement models, factor analysis, method variance, construct validation procedures, measurement design

‘‘ .. . [M]easurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules’’ (Stevens, 1946, p. 677), and it has long been recognized that the act of measuring some phenomenon can affect the phenomenon under study itself (Heisenberg, 1927; Messiah, 1961 1962), ‘‘by changing the underlying construct of interest or by distorting the mea- surement process’’ (Spector, 2006, p. 222). Depending on the particular approach, apparatus, proce- dure, or method that is used to assign numbers to objects or events, the effect that the act of measurement has upon the objects or events being measured may differ, and it is in this sense that

‘‘measurement method effect’’ is defined.

Measurement method (or more simply ‘‘method’’) effects are widely believed to pose threats to research in the organizational and behavioral sciences, especially when multiple study variables are measured using the same method. For example, according to Spector (2006) ‘‘[i]t is widely believed that relationships between variables measured with the same method will be inflated due to the

1 The University of Georgia

Corresponding Author:

Charles E. Lance, Department of Psychology, The University of Georgia, Athens, GA 30602-3013

Email: clance@uga.edu

action of common method variance’’ (p. 221). In addition, Podsakoff, MacKenzie, Lee, and Podsak- off (2003) argued that ‘‘[m]ost researchers agree that common method variance .. . is a potential problem in behavioral research’’ (p. 879). As a third example and as part of his outgoing comments as editor of the Journal of Applied Psychology, J. P. Campbell (1982) wrote:

With perhaps one or two exceptions there has been very little opportunity to exercise any professional biases (I think). 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 vari- ables that are measured independently of the questionnaire, I am biased against the study and believe that it contributes very little. Many people share this bias. (p. 692).

Thus, there appears to be a common belief, one might even say an ‘‘urban legend’’ (Lance & Vandenberg, 2009; Vandenberg, 2006) that correlations among variables that are measured using the same measurement method are distorted (typically inflated) as compared to correlations among the constructs themselves. Surprisingly, to our knowledge the idea that observed monomethod cor- relations are inflated as compared to correlations among the traits themselves has been tested only once previously (Lance & Sloan, 1993). The first purpose of this study was to provide a broader test of this idea using multiple published data sets.

The second purpose of this study was to estimate the pervasiveness and magnitude of method and common method effects in organizational research measures. This has been done before (Buckley, Cote, & Comstock, 1990; Cote & Buckley, 1987; Crampton & Wagner, 1994; Doty & Glick, 1998; Spector, 1987; Williams, Cote, & Buckley, 1989). However, we will argue that most of these pre- vious efforts suffered from one of two limitations that we intend to overcome. First, one approach to estimating common methods effects has been to compare the magnitudes of monomethod correla- tions (correlations between variables using the same method) with those of heteromethod correla- tions (correlations between variables using different methods). The general rationale of this approach is that the difference between monomethod and heteromethod correlations calibrates the extent of covariance distortion due to method effects. We will show later that this is generally incor- rect. We also argue below that recent research, which has shown that some measurement facets that once were considered mere measurement methods should actually be more properly regarded as comprising aspects of domain-specific substantive nomological networks (Cronbach & Meehl,

1955), and not merely alternative procedures for assigning numbers to objects or events (Stevens,

1946). In discussing these measurement facets, we will argue their importance and the need to study them in their own right, but that they should be excluded from consideration as mere measurement methods. Therefore, we aim to provide what we feel are more accurate estimates of measurement method and common method effects per se than have previous attempts to do so. We now turn to more detailed discussions of these two goals.

Are Monomethod Correlations Distorted (Usually Inflated) Compared to Correlations Among Corresponding Constructs?

Figure 1 illustrates two possible threats posed by method effects. The first possible threat is compro- mised construct validity due to contamination of observed measures by method effects:

Xijk ¼ lTij Ti þ lMij Mj þ Eijk ð1Þ

where Xijk is the kth realization of the ith person’s characteristic or trait (Ti) that is being measured by the jth method (Mj), and Eijk represents nonsystematic measurement error. For convenience, we

φ

*T T*

*i i* '

Ti Ti’

Mj

λ

λ

*T*

*ij*

*ij i* ' *j*

λ

λ

*i* ' *j*

*T*

Xijk Xi’jk

*M*

*M*

Eijk Ei’jk

Figure 1. Possible threats posed by method effects.

assume that E(Ti) ¼ E(Mj) ¼ E(Eik) ¼ 0 and that Xijk, Ti, and Mj have unit variances so that the ls effectively represent standardized regression coefficients (or factor loadings, see below). We also assume that E(Ti,Mj) ¼ E(Ti,Eijk) ¼ E(Mj,Eijk) ¼ E(Eijk,Eijk’) ¼ 0. Under these assumptions:

s2 2 2 2

Eijk

Xijk ¼ lTi þ lMj þ sEijk : ð2Þ

That is, variance in observed scores (s2

Xijk

Ti

) is a composite of true score variance (l2 ), method var-

iance (l2

Mj

), and nonsystematic error variance (s2

). Thus, the construct validity of X is compro-

mised to the extent that it reflects the influence of the measurement method relative to that of the focal trait, and it has been estimated that somewhere from 22% (Buckley et al., 1990) to 32% (Doty

& Glick, 1998) to 72% (Mount, Judge, Scullen, Sytsma, & Hezlett, 1998) of the variance in mea- sures used in organizational research is attributable to method variance (see also Cote & Buckley,

1987; Williams et al., 1989). However, the presence of method variance does not, by itself, neces- sarily pose a threat to research findings if, for example, method variance is not shared across mea- sures of different constructs.

A second possible threat of method effects is that of covariance distortion. As is illustrated in Figure 2, if two different traits (Ti and Ti0 ) are measured using the same measurement method, the observed score correlation reflects not just the correlation among the traits (FTi Ti0 ) but also the com- mon causal effect of the measurement method (lMij lMi0 j ). It is this latter effect that often leads to the attribution that measurement of different variables by the same measurement method leads to

inflated estimates of true score correlations, whereas the use of different methods avoids this covar- iance distortion and, in fact, if both lMij and lMi0 j are similarly signed and nonnegligible, an

φMjMj’

φTiTi

Mj

Ti Ti’

Mj’

λMij

λTij

λMi’j λ

Tij’ λTi’j λMij’ λTi’j’ λMi’j’

TiMj Ti’Mj TiMj’ Ti’Mj’

Eijk Eijk Eijk Eijk

Figure 2. CTCM CFA parameterization of MTMM data. CTCM ¼ correlated trait-correlated method; CFA ¼

confirmatory factor analysis.

inflationary effect will incur for the observed correlation rXij Xi0 j . However, this is not the whole story. rXij Xi0 j is also attenuated by the effects lTi and lTi0 , which are tantamount to the reliability indexes (or the square roots of the variables’ respective reliabilities in the absence of method effects; Crocker & Algina, 1986) for Xij and Xi’j:

rXij Xi0 j ¼ lTij lTi0 j fTi Ti0 þ lMij lMi0 j ð3Þ

Thus, monomethod correlations (or heterotrait monomethod [HTMM] correlations in multi- trait multimethod [MTMM] nomeclature, D. T. Campbell & Fiske, 1959) are simultaneously atte- nuated due to unreliability (to the extent that the lTi s are < 1.0), and distorted (usually inflated) due to common method effects (lMij lMi0 j ).

In one study that assessed these effects, Lance and Sloan (1993) conducted a confirmatory factor

analysis (CFA) of MTMM data consisting of a number of life facet satisfaction constructs (e.g., satisfaction with work, family, income, etc.) as measured by three different response scale formats. Using a correlated trait correlated method (CTCM) parameterization (Lance, Noble, & Scullen,

2002), Lance and Sloan found that the mean observed monomethod correlation (.30) was nearly identical to that of the mean satisfaction latent variable correlation (.29), indicating that the inflation- ary effect of common method variance almost exactly offset the attenuating effects of measurement error. These findings suggest that the widespread belief that monomethod correlations are inflated is supported by a kernel of truth (Vandenberg, 2006; i.e., there was an inflationary effect of common method effects) but is also part urban legend because the attenuating effect of unreliability offset this inflationary effect so that observed correlations were actually not higher than estimated correlations among the traits themselves. The first purpose of this study was to extend this research to the broader literature that has used MTMM matrices.

What are the Magnitudes of Method and Common Method Effects?

As we mentioned earlier, there have been several previous attempts to answer this question, and the answer seems to depend on which source one consults and how these effects are estimated. Some (e.g., Buckley et al., 1990; Cote & Buckley, 1987; Doty & Glick, 1998; Williams et al., 1989) have claimed that method effects are widespread, robust, and pose the threats discussed earlier to orga- nizational and social science research whereas others (e.g., Crampton & Wagner, 1994; Spector,

1987, 2006) have claimed that the potential threats posed by method effects are minimal. We hope

to clarify this controversy in two ways and provide more accurate estimates of the prevalence of method effects.

Do Not Compare Monomethod and Heteromethod Correlations to Assess Method Effects

One approach to gauging the presence of method effects has been to compare the magnitudes of monomethod correlations with those of heteromethod correlations under the rationale that mono- method correlations should be inflated by common method bias whereas the heteromethod correla- tions should not. However, this is generally not true. We show why in reference to Figure 2 (of which Figure 1 is a subset), which shows a general (but simplified) representation of relationships between multiple traits as measured by multiple methods in a MTMM matrix (D. T. Campbell & Fiske,

1959). As we showed earlier, a HTMM correlation can be decomposed from Figure 1 or the more general Figure 2 as

rTi Mj ;Ti0 Mj ¼ lTij lTi0 j fTi Ti0 þ lMij lMi0 j : ð4Þ

A HTMM correlation reflects the influence of possibly correlated traits (attenuated by measurement error) plus the (usually) inflationary effects of common method effects. However, covariance distortion effects are not limited to monomethod correlations. As is shown in Figure 2, monotrait heteromethod (MTHM) and heterotrait heteromethod (HTHM) correlations are also potentially subject to covariance distortion, because they reflect the influences of potentially correlated methods:

MTHM ¼ rTi Mj ;Ti Mj0 ¼ lTij lTij0 þ lMij lMij0 fMj Mj0 ð5Þ

HTHM ¼ rTi Mj ;Ti0 Mj0 ¼ lTij lTi0 j0 fTi Ti0 þ lMij lMi0 j0 fMj Mj0 ð6Þ

Thus, contrary to conventional wisdom, the so-called ‘‘convergent validities’’ (D. T. Campbell & Fiske, 1959; Woehr & Arthur, 2003, p. 242) reflect the influence not only of the common trait (lTij lTij0 ) but also potentially of correlated methods (lMij lMij0 fMj Mj0 ). Even the HTHM correlations,

which logically have nothing in common (these correlations reflect correlations between different

traits as measured by different methods), still potentially reflect the influences of both correlated traits (lTij lTi0 j0 fTi Ti0 ) and correlated methods (lMij lMi0 j0 fMj Mj0 ). Thus, to the extent that methods’ effects on observed measures are non-negligible and homogeneously signed (e.g., all positive) and method factors correlate positively, covariance inflation will incur even for heteromethod correla- tions, including the so-called convergent validities (MTHM correlations). In fact, the so-called

‘‘convergent validities’’ (Woehr & Arthur, 2003, p. 242) can be reasonably high even when there is no convergent validity (effects of the common trait are 0, i.e., lTi ¼ lTi0 ¼ 0) if the effects of rea- sonably highly correlated methods are reasonably strong (i.e., the MTHM correlation is exclusively a function of the product lMij lMij0 fMj Mj0 ). Of course, if the methods are uncorrelated, no covariance distortion will incur, even in the presence of substantial method variance, and if the methods are negatively correlated, additional attenuation beyond that due to the measures’ unreliability will

incur. This analysis points to the single case in which comparisons between monomethod and het- eromethod correlations provide an appropriate indication of covariance distortion: methods must be uncorrelated. Are they? Not according to Doty and Glick (1998) who found that the mean estimated methods correlation ðfMj Mj0 Þ in the studies reviewed was .34, or Buckley et al. (1990) whose mean estimated fMj Mj0 was .37, or Cote and Buckley (1987) whose mean estimated fMj Mj0 was .48 or Wil- liams et al. (1989) whose mean estimated fMj Mj0 was .63. Different raters or rater sources are some- times treated as different methods (however, see below) and Hoffman, Lance, Bynum, and Gentry (in press) recently estimated the correlation between boss and peer rating source latent variables to be .86. Thus, the assumption that methods are uncorrelated is not generally supported and neither is the practice of comparing monomethod to heteromethod correlations to assess method effects. Results from studies that have found small differences between monomethod and heteromethod cor- relations are ambiguous because these results could indicate that either (a) method effects are small in both Equations 4 (lMij lMi0 j ) and 6 (lMij lMij0 fMj Mj0 ), or (b) method effects were present in both cases and method correlations (fMj Mj0 ) were non-negligible, in the absence of direct estimates for the lMs and fMj Mj0 s. As such, comparisons between monomethod and heteromethod correlations are ambiguous at best and are potentially quite misleading.

Not all Alleged Methods Should Really be Considered Mere Measurement Methods

D.T. Campbell and Fiske’s (1959) design has been applied to a wide variety of trait domains and a number of different alleged measurement methods, including different raters (or rater sources), tests, response formats, exercises in assessment centers (ACS), measurement occasions, apparatuses, and so on (Lance, Baranik, Lau, & Scharlau, 2009). Appropriately, Spector (2006) questioned ‘‘What is a method?’’ (p. 227). That is, what is it that qualifies each of these measurement facets as a measure- ment method facet as compared to the trait measurement facet? In their recent review of MTMM literature, Lance et al. (2009) suggested that researchers sometimes have been rather cavalier in their treatment of method facets under a default assumption that ‘‘if it ain’t trait it must be method’’ (p.

339), meaning that if one has a three-facet measurement design where one facet represents the objects of measurement (e.g., research participants), a second facet that represents the focal con- structs (i.e., the traits of interest, such as personality or job performance dimensions), then the third facet (e.g., raters, AC exercises, response formats) that is not the trait facet must therefore be a mea- surement method facet. One danger in this default assumption is that method effects and common method variance are usually assumed to reflect sources of unwanted contamination and bias that should be minimized (Avolio, Yammarino, & Bass, 1991). One (typical) example of this attribution is the claim by Bagozzi and Yi (1990) that ‘‘method variance can bias results when researchers investigate relations among constructs with the common method ... method variance produces a potential threat to the validity of empirical findings’’ (p. 547). As another example, Ostroff, Kinicki, and Clark (2002) referred to the ‘‘ ... insidious effects of method variance and method bias’’ (p.

366). However, some measurement facets that were once thought of as mere measurement method facets that contributed unwanted contaminating bias in the measurement process are now better understood as representing substantive theoretical constructs in their own right that affect organiza- tional outcomes (see Lance et al., 2009 for a review).

As one example, researchers in the area of multisource performance rating (where performance dimensions were treated as traits and raters or rater sources as methods) long considered individual rater and rater source effects as representing artifactual and contaminating rater biases such as halo error (e.g., Conway & Huffcutt, 1997; King, Hunter, & Schmidt, 1980; Viswesvaran, Ones, & Schmidt, 1996; Wherry & Bartlett, 1982). However, more recent research has demonstrated

empirically that these rater and source effects are more appropriately interpreted as sources of per- formance true-score-related variance (Hoffman & Woehr, 2009; Lance, Baxter, & Mahan, 2006; Lance, Hoffman, Gentry & Baranik, 2008; Lance, Teachout & Donnelly, 1992). As Borman (1974) suggested over 35 years ago, rater and source effects represent different aspects of the criter- ion construct space that are captured uniquely by the different sources. The fact that rater/source effects dominate over dimension effects on multisource ratings (see Hoffman et al., in press; Lance et al., 2008; Mount et al., 1998; Scullen, Mount, & Goff, 2000) led researchers for years to believe that multisource ratings were biased and had low interrater reliability (e.g., Viswesvaran et al.,

1996), when, in fact, raters were doing just what practitioners had envisioned in the first place–– representing their different, yet somewhat complementary perspectives on ratee performance (e.g., London & Tornow, 1998). Thus, different raters and rater sources should not be considered mere measurement methods.

A second example comes from MTMM-related AC construct validity research where AC dimen- sions were historically treated as traits and AC exercises as methods (e.g., Arthur, Woehr, & Maldegan, 2000; Lievens & Conway, 2001; Woehr & Arthur, 2003). Early on in research on the construct validity of ACs, Sackett and Dreher (1982) found that exercise (qua method) effects dom- inated over dimension (qua trait) effects in factor analyses of AC ratings, referring to these as a set of

‘‘troubling findings’’ (p. 401). Now, after more than 25 years of attempts to ‘‘fix’’ ACs so that they do not exhibit these allegedly unwanted exercise (qua ‘‘method’’) effects, AC exercise effects still dominate over dimension effects (see Lance, 2008a; Lance, Lambert, Gewin, Lievens, & Conway,

2004). Now, however, these exercise effects are more properly understood as representing cross- situational specificity in AC performance not method bias (Lance, 2008a, 2008b; Lance et al.,

2000). As such, AC construct validity research represents another area where the default assumption that the third facet in a measurement design must be a method facet was invoked, and had the effect of sending researchers on a 25-year long detour to fix an organizational intervention that was not broken in the first place. AC exercises never should have been and should not now be considered mere measurement methods.

Other measurement facets that once were thought of as mere methods but are now better under- stood as having substantively relevant effects on theoretically important constructs include measure- ment occasions, where occasion factors are now better understood as representing state-related aspects of the traits being measured (e.g., Kenny & Zautra, 2001; Schermelleh-Engel, Keith, Moosbrugger, & Hodapp, 2004), and positive–negative item wording effects that have been shown to be related predictably to personality traits such as neuroticism emotional stability (e.g., DiStefano & Motl, 2006; Quilty, Oakman, & Risko, 2006).

Lance et al. (2009) provided a prototypical multidimensional measurement system that was pat- terned after Cattell’s (1966) ‘‘basic data relation matrix’’ or ‘‘data box.’’ Lance et al. proposed that most, if not all, MTMM-related designs seen in the organizational research literature could be located as a three-dimensional subspace within the measurement facets of:

(a) persons (or groups of persons, or collectivities of groups of persons who may be the object of study); (b) focal constructs that constitute the relevant characteristics of the entities studied; (c) occasions, or temporal replications of measurement; (d) different situations in which measurement may occur; (e) observers or recorders of entities’ behavior; and (f) response modalities/formats (p. 354).

For example, the multisource ratings data discussed earlier combines the persons (i.e., ratees), focal constructs (i.e., rating dimensions), and observers (i.e., raters) facets, and the AC data discussed ear- lier combines the persons (i.e., candidates), focal constructs (i.e., dimensions), and situations (i.e., exercises) facets. Lance et al. (2009) encouraged researchers to locate their research design within this larger multidimensional data array to help determine whether some measurement facet

represents a mere method or whether it represents something more substantively meaningful, and to avoid the default assumption that ‘‘if it ain’t trait it must be method’’ (p. 339). Lance et al. suggested that:

If a particular measurement facet truly represents alternative approaches to assigning numbers to obser- vations to represent individuals’ standing on latent constructs independent of substantive content related to other latent constructs, the facet might be reasonably viewed as representing alternative measurement methods (p. 353).

We followed this advice in the selection of data sets for reanalysis in the current study.

Summary

In summary, the current study’s objectives were to (a) determine the veracity of the urban legend that monomethod correlations are inflated by common method variance as compared to the correla- tions among the traits themselves and (b) estimate trait and method variance and covariance com- ponents for MTMM data in which the alleged methods truly represent mere variations in measurement approaches.

Method

Selection of Studies

We identified relevant articles using the PsychInfo database by conducting a citation search on the

D. T. Campbell and Fiske (1959) article, using the subject keywords ‘‘multitrait multimethod’’ and

‘‘MTMM’’ and by locating the primary studies cited by Cote and Buckley (1987), Doty and Glick (1998), Williams et al. (1989), and Buckley et al. (1990). A total of 1,711 nonredundant articles were identified. We limited our review to sources that were in English and that were available online. This resulted in the elimination of 17 and 94 sources, respectively (most of the latter were older journal articles or book chapters). Another 1,226 articles cited the article of D. T. Campbell and Fiske (1959) but for reasons other than the conduct of an MTMM study. For example, many of these citations were to general references to measures’ convergent, discriminant, or construct validity (e.g., D’Mello & Graeser, 2009; Franic & Bothe, 2008). Another 80 studies apparently did conduct MTMM analyses, up to the point of presenting descriptive information (e.g., mean HTMM, MTHM, and HTHM correlations) but did not present the actual MTMM matrix (e.g., Choi, 2004; Lievens, Chasteen, Day, & Christiansen, 2006). Another 138 articles were not included because they were theoretical, primarily methodologically oriented, and/or nonempirical. Examples of these include the demonstration of MTMM approach for longitudinal research by Cole (2006), and the critique of the correlated uniqueness (CU) model by Lance et al. (2002) for MTMM data. An additional

80 studies were eliminated because the alleged ‘‘methods’’ actually represented substantive mea- surement facets: (a) 55 studies were excluded because they used raters as methods (e.g., multisource performance ratings, parent, teacher, peer, and self ratings of school performance, etc.), (b) 14 stud- ies used AC exercises as methods, and (c) 11 studies used multiple measurement occasions as meth- ods. Another 16 studies were eliminated because they were underidentified in a CTCM CFA parameterization of MTMM data (see Lance et al., 2002).1 Finally, 42 studies provided otherwise usable MTMM matrices but were not retained because they resulted in nonconvergent and/or inad- missible CFA solutions (see below). Unfortunately, this stringent elimination process resulted in the inclusion of only the 18 studies described below. Even so, these studies represent a diverse set of constructs ranging from various aspects of personality, organizational reactions and perceptions, attitudes, work performance, and vocational interests and thus support generalizable findings.

Analyses

Each candidate MTMM matrix was input to LISREL 8.8 (Jo¨ reskog & So¨ rbom, 1993) for a CTCM CFA parameterization. We chose the CTCM model for analysis because it is theoretically the most faithful to the design of the MTMM matrix (Lance et al., 2002). We did not attempt CU parameter- izations of method effects because the CU model requires the assumption that methods are uncor- related, an assumption that we documented earlier is untenable. Furthermore, the CU model is known to produce upwardly biased estimates of trait factor loadings and correlations, despite its advantage over the CTCM model in returning admissible solutions (Conway, Lievens, Scullen, & Lance, 2004; Lance et al., 2002). We also did not test trait- or method-only models as we wished to estimate both trait and method effects on measures simultaneously. We also did not test uncorre- lated methods model because (a) previous studies have documented that methods are routinely cor- related, (b) an uncorrelated methods model also produces upwardly biased estimates of trait factor loadings and intercorrelations as does the CU model, and (c) an uncorrelated methods model is a special case of the correlated methods model that is routinely estimable (method correlations can be estimated at any value, including 0) under the CTCM model. We allowed a maximum of

1,000 iterations and determined that a solution was admissible if all uniquenesses were non- negative and factor loadings and correlations were <1.00 in absolute value. In the event that a solu- tion was found to be inadmissible, we attempted a mathematically equivalent parameterization of the CTCM model presented by Rindskopf (1983):

2 FTT0

L

32 0 3

T

0

7

MTMM ¼ ½ LT LM LU1=2 4

0 FMM0

0 0 I

5

4

6 LM

LU 1=2

0

5 ð7Þ

where, assuming that traits and methods are crossed (though they need not be), LT and LM are the

T\*M T and T\*M M matrices of trait and method factor loadings, respectively, LU 1=2 is the T\*M

 T\*M diagonal matrix containing square roots of the uniquenesses, FTT’ (T T) and FMM’ (M

M) contain correlations among traits and methods, respectively, and I is a T\*M T\*M identity

matrix. Because uniquenesses are calculated as the squares of the elements in LU 1=2 , even negative values are considered proper.

Results

Table 1 shows the number of traits and methods in, and the overall goodness-of-fit indices for each of the MTMM matrices that returned convergent and admissible solutions. Overall, model fit was good to excellent for the CTCM model except for the Kothandapani (1971) and Yang, Lance, and Hui (2006) studies, the latter of which reported results of a relatively sparsely parameterized model.

Table 2 provides a brief description of each study’s traits and methods and each study’s mean correlations and trait and method factor loadings and correlations. The traits measured included var- ious personality dimensions, attitudinal dimensions, and work-related variables. Methods were dif- ferent response formats for the same items (e.g., Stapel, Likert, Semantic Differential formats, Arora, 1982) or different scales/subscales designed to measure the same traits (e.g., the job diagnos- tic survey and the job rating form to measure Hackman & Oldham’s [1976] core job characteristics, Birnbaum, Farh, & Wong, 1986). HTMM correlations were highest for job-related variables (M ¼

.42), followed by attitudes (M ¼ .35) and personality dimensions (M ¼ .20; F(2,14) ¼ 2.84, p < .10), MTHM correlations were higher for scale format methods (M ¼ .61) than for different tests as meth- ods (M ¼ .47; F(1,16) ¼ 3.43, p < .10) and trait factor loadings were higher for scale format methods (M ¼ .72) than for different tests as methods (M ¼ .58; F(1,16) ¼ 6.98, p < .05).

Table 1. Number of Traits and Methods and Overall Goodness-of-fit Indices for MTMM Matrices

|  |  |  |  |
| --- | --- | --- | --- |
| Study | Traits | Methods | df |

|  |  |  |  |
| --- | --- | --- | --- |
| w2 SRMSR | RMSEA | TLI | CFI |
| Allen and Hamsher (1974) | 3 | 3 | 12 | 7.23 | .025 | <.001 | 1.01 | 1.00 |
| Arora (1982) | 3 | 3 | 12 | 30.83 | .066 | .091 | .95 | .98 |
| Birnbaum, Farh, and Wong (1986) | 5 | 2 | 14 | 15.91 | .039 | .031 | .98 | .99 |
| Flamer (1983) Study 1 | 3 | 3 | 12 | 8.91 | .022 | <.001 | 1.01 | 1.00 |
| Flamer (1983) Study 2 | 3 | 3 | 12 | 20.96 | .052 | .059 | .96 | .99 |
| Hewitt, Foxcroft, and MacDonald (2004) | 3 | 6 | 99 | 191.10 | .018 | .019 | .99 | .99 |
| Johnson, Smith, and Tucker (1982) | 5 | 2 | 14 | 24.75 | .057 | .086 | .95 | .98 |
| Keaveny and McGann (1975) | 13 | 2 | 194 | 351.42 | .040 | .059 | .98 | .99 |
| Kothandapani (1971) | 3 | 4 | 33 | 240.13 | .066 | .120 | .89 | .95 |
| Lance, Noble, and Scullen (2002) | 3 | 3 | 12 | 6.66 | .021 | <.001 | 1.06 | 1.00 |
| Mosher (1966) | 3 | 3 | 12 | 6.04 | .016 | <.001 | 1.02 | 1.00 |
| Palmer and Loveland (2004) | 5 | 2 | 14 | 23.27 | .042 | .066 | .94 | .98 |
| Pitoniak, Sireci, and Luecht (2002) | 4 | 3 | 24 | 1576.76 | .012 | .041 | .99 | 1.00 |
| Rose and Andiappan (1978) | 4 | 4 | 76 | 95.46 | .074 | .058 | .93 | .98 |
| Seligson, Huebner, and Valois (2003) Study 1 | 5 | 2 | 14 | 12.81 | .021 | <.001 | 1.01 | 1.00 |
| Seligson et al. (2003) Study 2 | 5 | 2 | 14 | 15.76 | .049 | .036 | .96 | 1.00 |
| Thompson and Berven (2002) | 4 | 2 | 5 | 5.38 | .016 | .014 | .99 | 1.00 |
| Yang, Lance, and Hui (2006) | 6 | 5 | 350 | 1899.56 | .130 | .076 | .92 | .94 |

Note: Traits ¼ number of traits; Methods ¼ number of methods; df ¼ models degrees of freedom; SRMSR ¼ standardized root mean squared residual; RMSEA ¼ root mean squared error of approximation; TLI ¼ Tucker-Lewis Index; CFI ¼ Comparative Fit Index.

Our study’s first objective was to test the idea that monomethod correlations are inflated over actual correlations among traits. The penultimate row in Table 2 shows that the mean HTMM correlation (.340) was actually slightly lower than the average estimated trait intercorrelation (.371). A return to Equation 3 explains why this result incurs: HTMM correlations are inflated by common method

2

variance by a factor of lMij lMi0 j ¼ :427

¼ :182 on the average, but the estimated trait intercorrela-

2

tion ¼ .371 is also attenuated due to measurement error by a factor of lTij lTi0 j ¼ .635

¼ .403 on the

average so that the attenuated true score correlation is reduced by .222 to .149. The HTMM that is

reproduced on the basis of CFA model parameter estimates, r^Ti Mj ;Ti0 Mj ¼ l^Tij l^Ti0 j f^ Ti T

i0

þl^Mij l^Mi0 j ¼

.6352\*.371 þ .4272 ¼ .332 is nearly identical to the mean HTMM estimated directly from the data

(further attesting to the goodness-of-fit of the CTCM model), and shows (as in the Lance & Sloan,

1993 article) that common method effects largely offset the attenuating effects of unreliability in monomethod correlations. Thus, there is a kernel of truth to the urban legend that common method effects inflate monomethod correlations, but it is a myth that monomethod correlations are larger than their true score counterparts and this is because of the attenuating effects of measurement error.

Results in Table 2 also serve to illustrate a point that we made earlier––differences between mono- method and heteromethod correlations should not be used to estimate common method effects. As we

2

just showed, the common method effect is estimated from results here as lMij lMi0 j ¼ .427

¼ .182 on

the average, whereas the difference between the mean HTMM and HTHM correlations is .340

.248 ¼ .092, 48% lower than that estimated from CFA results. This is because 39% of the .248 mean

HTHM correlation is itself due to correlated method variance, or lMij lMi0 j0 fMj Mj0 ¼ .427 \*.520 ¼

2

.095 from Table 2. Thus, to the extent that methods are correlated, as it appears from previous reviews and Table 2 that they routinely are, differences between HTMM and HTHM correlations will return inaccurate underestimates of common method variance. Note that two reviews that con- cluded that covariance distortion of common method effects is minimal (Crampton & Wagner, 1994; Spector, 1987) used this approach.

445

|  |  |  |
| --- | --- | --- |
| Table 2. Study Descriptions | and CFA Results |  |
| Study | Traits | Methods | MTHM | HTMM | HTHM | TL | ML | TR | MR |
| Allen and Hamser (1974) | 3 dimensions of emotionality | Marlowe-Crowne | .608 | .522 | .367 | .588 | .622 | .387 | .753 |
|  |  | Rotter Scale |  |  |  |  |  |  |  |
|  |  | Mosher Scale |  |  |  |  |  |  |  |
| Arora (1982) | Situational, enduring and response | Stapel, Likert, Semantic | .669 | .339 | .226 | .769 | .458 | .183 | .433 |
|  | involvement | Differential |  |  |  |  |  |  |  |
| Birnbaum, Farh, and Wong(1986) | Core job dimensions based onHackman and Oldham | Job diagnostic surveyJob Rating Form | .330 | .390 | .180 | .590 | .350 | .500 |  .200 |
| Flamer (1983) Study 1 | Attitudes towards discipline, the lawand mathematics | LikertThurstone | .540 | .057 | .044 | .587 | .366 |  .103 | .877 |
|  |  | Semantic differential |  |  |  |  |  |  |  |
| Flamer (1983) Study 2 | Attitudes towards discipline, the law | Likert | .534 | .093 | .075 | .683 | .302 | .063 | .700 |
|  | and mathematics | Thurstone |  |  |  |  |  |  |  |
|  |  | Semantic differential |  |  |  |  |  |  |  |
| Hewitt, Foxcroft, and | Internality, stability, & globality | 6 items from the attributional | .190 | .290 | .100 | .340 | .500 | .280 | .270 |
| MacDonald (2004) |  | style questionnaire |  |  |  |  |  |  |  |
| Johnson Smith and Tucker | 5 Job description dimensions | 2 Job Descriptive Index scale | .660 | .393 | .3105 | .758 | .468 | .332 | .630 |
| (1982) |  | formats |  |  |  |  |  |  |  |
| Keaveny and McGann (1975) | 13 categories of teaching | Graphic rating scale | .635 | .485 | .431 | .799 | .238 | .659 | .060 |
|  | performance | Behavioral expectation scale |  |  |  |  |  |  |  |
| Kothandapani (1971) | Feeling, Belief, and Intention | Thurstone | .481 | .449 | .179 | .643 | .603 | .297 | .195 |
|  |  | Likert |  |  |  |  |  |  |  |
|  |  | Guttman |  |  |  |  |  |  |  |
|  |  | Guilford |  |  |  |  |  |  |  |
| Lance, Noble, and Scullen | Attitudes towards diversity | Structured Interview | .160 | .320 | .150 | .410 | .370 | .790 | .230 |
| (2002) |  | In-Basket |  |  |  |  |  |  |  |
|  |  | Biodata Form |  |  |  |  |  |  |  |
| Mosher (1966) | Sex guilt, Hostile guilt; Mortality/ | Forced Choice | .770 | .520 | .490 | .780 | .360 | .660 | .860 |
|  | Conscience guilt | True False |  |  |  |  |  |  |  |
|  |  | Incomplete Sentence |  |  |  |  |  |  |  |

(continued)

446

|  |  |
| --- | --- |
| Table 2 (continued) |  |
| Study | Traits | Methods | MTHM | HTMM | HTHM | TL | ML | TR | MR |
| Palmer and Loveland (2004) | Big five personality traits | 2 Big Five Scales: | .710 | .171 | .135 | .697 | .331 | .212 | .860 |
|  |  | Goldberg, Mini-Marker |  |  |  |  |  |  |  |
| Pitoniak, Sireci, and Luecht | 4 factors of the the uniform CPA | 3 different item formats | .610 | .540 | .670 | .600 | .520 | .710 | .930 |
| (2002) | Examination |  |  |  |  |  |  |  |  |
| Rose and Andiappan (1978) | 4 Work Performance Dimensions | 4 Sex-role context | .511 | .419 | .281 | .669 | .479 | .490 | .447 |
| Seligson, Huebner, and Valois | 5 Satisfaction targets | 2 Life Satisfaction Scales: MSLSS, | .636 | .433 | .389 | .676 | .509 | .377 | .740 |
| (2003) Matrix 1 |  | BMSLSS |  |  |  |  |  |  |  |
| Seligson et al. (2003) Matrix 2 | 5 Satisfaction targets | 2 Life Satisfaction Scales: MSLSS, | .534 | .368 | .240 | .695 | .451 | .349 | .320 |
|  |  | BMSLSS |  |  |  |  |  |  |  |
| Thompson and Berven (2002) | Horizontal individualism (HI) | 2 individual/collectivism scales: | .448 | .097 | .058 | .511 | .319 | .187 | .720 |
|  | Vertical individualism (VI) | FFIC, ICVAQ |  |  |  |  |  |  |  |
|  | Horizontal collectivists (HC) |  |  |  |  |  |  |  |  |
|  | Vertical collectivists (VC) |  |  |  |  |  |  |  |  |
| Yang, Lance, and Hui (2006) | 6 Factors of the SDS representing | Five Self-directed search (SDS) | .480 | .240 | .140 | .630 | .440 | .310 | .540 |
|  | vocational interests | subtests |  |  |  |  |  |  |  |
| Mean |  |  | .528 | .340 | .248 | .635 | .427 | .371 | .520 |
| SD |  |  | .165 | .154 | .168 | .122 | .104 | .233 | .321 |

Note. MTHM ¼ mean monotrait-heteromethod correlation; HTMM ¼ mean heterotrait-monomethod correlation; HTHM ¼ mean heterotrait-heteromethod correlation; TL ¼ mean CFA-estimated trait factor loading; ML ¼ mean CFA-estimated method factor loading; TR ¼ mean CFA-estimated trait factor correlation; MR ¼ mean CFA-estimated method factor correlation.

Table 3. Comparative Estimates of Trait and Method Variance and Correlations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Study | % Trait variance | % Method variance | Trait correlation | Method correlation |
| Cote and Buckley (1987) | .42 | .26 | .67 | .48 |
| Doty and Glick (1998) | .46 | .32 | .26 | .34 |
| Williams, Cote, and Buckley (1989) | .51 | .27 | NR | .63 |
| Buckley, Cote, and Comstock (1990) | .42 | .22 | .51 | .37 |
| Present study | .40 | .18 | .37 | .52 |
| Raters as Methods: |  |  |  |  |
| Mount, Judge, Scullen, Sytsma, and Hezlett(1998) | .09 | .72 |  .02 | .15 |
| Scullen, Mount, and Goff (2000) | .23 | .65 | NR | NR |
| Hoffman, Lance, Bynum, and Gentry | .10 | .77 | .13 | .45 |
| (in press) |  |  |  |  |
| AC exercises as methods: |  |  |  |  |
| Lievens and Conway (2001) | .34 | .34 | .71 | NR |
| Lance, Lambert, Gewin, Lievens, and | .14 | .52 | NR | .42 |
| Conway (2004) |  |  |  |  |
| Bowler and Woehr (2006) | .22 | .34 | .79 | .10 |

Finally, the MTHM correlation reproduced from CFA results

þl^Mij l^Mij0 f^ Mj M

r^Ti Mj ;Ti Mj0 ¼ l^Tij l^Tij0

j0 ¼ .635

2

2

þ .427 \*.520 ¼ .498 was only slightly lower than that estimated directly

from the data (.528). These results illustrate the point made earlier that nearly 20% of the mean so- called convergent validity correlation (Woehr & Arthur, 2003) comes not from convergent validity (lTij lTij0 ) but from correlated method effects (lMij lMij0 fMj Mj0 ).

The second goal of this article was to provide more accurate estimates of the pervasiveness of

method effects in organizational research. Results from Table 2 are summarized in the top panel of Table 3 along with previous studies’ estimates of trait and method variance components and cor- relations. Our estimate of the proportion of variance attributable to method variance (18%) was somewhat lower than previous studies’ estimates. Post hoc, we speculated that the reason that our estimate of the mean proportion of method variance was lower than previous reviews was that we excluded a number of studies that represented substantively meaningful measurement facets (i.e., raters, AC exercises, and occasions as methods). Cote and Buckley (1987) did not list the stud- ies included in their review, but the other three reviews listed in the top portion of Table 3 did. The studies reviewed by Doty and Glick (1998) used a variety of different measurement methods. We compared the proportions of method variance reported by Doty and Glick for those studies that used raters and AC exercises to those studies that used different tests and response formats and found that the former yielded significantly larger proportions of method variance, on the average (M ¼ 43%) as compared to the latter (23%, F(1,26) ¼ 14.23, p < .001), supporting our idea. The studies reviewed by Williams et al. (1989) all used different tests or response formats as methods. The studies reviewed by Buckley et al. (1990) used a variety of methods and the mean estimated method var- iance for raters as methods was higher (24%) than for different tests or response formats (21%), but this difference was nonsignificant. Thus, there is some support for the idea that by excluding sub- stantive measurement facets, we obtained a lower and perhaps more accurate estimate of the preva- lence of method variance in organizational measures.

We present summary results from six other studies in the lower panel of Table 3 that were

conducted after the major reviews of method variance in organizational research cited in the upper portion of the Table to provide some additional context on this interpretation. Here we cite three

large-scale primary studies’ summary results that used somewhat different CFA parameterizations to estimate rater, source, and dimension variance components in multisource performance ratings (Hoffman et al., in press; Mount et al., 1998; Scullen et al., 2000). Results show that if raters were (inappropriately) considered methods, then strong and pervasive method effects would be implicated for multisource ratings. In addition, Bowler and Woehr (2006), Lance et al. (2004), and Lievens and Conway (2001) reported three large-scale reanalyses of AC data, again using somewhat different analytic approaches. Here again, if AC exercises were (inappropriately) considered methods, strong method effects would be implicated. Thus, we interpret these patterns of results as indicating that if substantively meaningful measurement facets are inappropriately classified as method factors, that (common) method variance will appear to be a far more pervasive phenomenon than it really is.

Discussion

Our two main findings from an updated review of MTMM studies can be summarized as follows: (a) there is a kernel of truth to the urban legend that common method effects inflate monomethod cor- relations but it is a myth that monomethod correlations are larger than correlations among the con- structs themselves, and this is because of the offsetting and attenuating effects of measurement error. Thus, monomethod correlations are generally not inflated as compared to their true score counter- parts, and (b) method variance occurs frequently, but when substantive measurement facets are excluded as alleged methods, common method variance does not appear to be as robust and threa- tening as many seem to think.

Due to the counterbalancing effects of common method effects and measurement unreliability, monomethod correlations were almost exactly equal to (and were actually slightly lower) than their estimated true score counterparts, on the average. This suggests that fear that ‘‘everything correlates to some extent with everything else’’ (Meehl, 1990a, p. 204) because of common method bias (the

‘‘crud factor,’’ Meehl, 1990a, 1990b) may not be so justified after all, despite being repeatedly

refueled and fanned by instructors, researchers, reviewers, and editors (e.g., J. P. Campbell,

1982). This also suggests that monomethod research may represent more accurate portrayal of rela- tions among constructs themselves than has been thought, and may even be more accurate than het- eromethod research because the latter are more likely to be attenuated as compared to their true score counterparts (see also Frese & Zapf, 1988).

These findings also have implications for statistical corrections for attenuation and common method effects. Returning to Equation 4, attempts to control statistically for common method effects by removing the lMij lMi0 j component from the HTMM correlations rTi Mj ;Ti0 Mj ¼ lTij lTi0 j fTi Ti0

þlMij lMi0 j , for example, using the partial correlation procedure by Lindell and Whitneys (2001)

could result in an underestimate of the correlation among the traits themselves (i.e., fTi Ti0 is attenu- ated by a factor equal to lTij lTi0 j ). Thus, if statistical controls are effected for common method effects, the resulting partial correlation should be corrected for attenuation with accurate estimates

of the respective variables’ reliabilities. Of course, correcting for attenuation due to unreliability without simultaneous correction for common method effects poses the reverse problem––the result- ing corrected correlation could likely be an overestimate of the true score counterpart. In addition, unfortunately, these complications are not limited to monomethod correlations, because as we showed earlier, HTHM correlations potentially reflect both the influences of correlated traits and correlated methods (HTHM ¼ lTij lTi0 j0 fTi Ti0 þ lMij lMi0 j0 fMj Mj0 ).

So what is one to do? We see five options. First, one can continue to ignore possible threats of

unreliability and common method effects. We suspect that this will continue to be a frequently adopted choice and, paradoxically, this actually may not be such a bad idea after all if the inflation- ary effects of common method effects do offset attenuating effects of measurement error in a

particular study. Second, one can adopt multiple indicator models and model either measured or unmeasured method factors as is discussed by Williams and Anderson (1994) and Podsakoff et al. (2003). Little is known about this approach, though research so far shows that modeling mea- sured method factors have surprisingly little effect on structural model parameter estimates (Wil- liams & Anderson, 1994). Third, one might adopt the partial correlation approach by Lindell and Whitney (2001) to adjust observed correlations for the effects of common methods variance. How- ever, little is known about the effectiveness of this approach either.

Developments presented in this article suggest a fourth approach, a simultaneous correction for

attenuation and method effects for monomethod correlations:

rXij Xi0 j lMij lMi0 j

r^Xij Xi0 j ¼

lTij

lTi0 j

ð8Þ

or for heteromethod correlations:

rXij Xi0 j0 lMij lMi0 j0 F Mj Mj0

r^Xij Xi0 j0 ¼

lTij

lTi0 j0

ð9Þ

where the rs represent the disattenuated and method variance-adjusted correlations, the rs are the corresponding observed correlations, lMs are the effects of measurement methods on observed mea- sures, FMj Mj0 is the correlation among different methods, and lTs are reliability indexes (square roots of reliability coefficients). A simultaneous multivariate correction could also be effected as:

C ¼ D 1=2 ðR MÞD 1=2 ð10Þ

where C is the matrix of corrected correlations, D is a diagonal matrix containing the variables’ reli- abilities, R is the original correlation matrix, and M is a square symmetric matrix containing correc- tions for method effects. For example, if variables A and B are measured using the same method (method 1), but variable C is measured using method 2, M would be constructed:

2 0 SYM 3

M ¼ 4

lA1 lB1 0

lA1 lC2 FM1 M2 lB1 lC2 FM1 M2 0

5: ð11Þ

Of course, this raises the question of where one would obtain estimates of the quantities needed to effect the corrections. Reliability estimates for most measures are readily available or computable. lMs can be estimated from a primary CFA of MTMM data. These quantities have also been esti- mated previously in many primary studies and are reported for a variety of different methods in table

4 by Doty and Glick (1998), table 2 by Williams et al. (1989), table 1 by Buckley et al. (1990), and the current study’s Table 2. FMjMj’s can also be estimated from a primary CFA of MTMM data or estimates from previous studies can be located in table 1 by Buckley et al. (1990) and the current study’s Table 2. In both cases prior estimates of similar methods’ effects and correlations can be used to effect the corrections in Equations 8–11.

Finally, one might conduct a CFA on MTMM data to confirm a measurement model and then conduct structural equation modeling on the trait intercorrelations, leaving the method loadings and correlations in place to control for method effects (see Arora, 1982 for an example of this approach using a CU parameterization of orthogonal method effects). Of course, this is the most direct approach to the control of method effects in a particular study but also has the obvious drawback of requiring multiple (different method) indicators for each modeled construct. This points to the first need for future research that we suggest: the latter four approaches to the modeling and control

of method effects need to be studied individually and in comparison to one another to determine their relative effectiveness.

One point that we have emphasized in this article is that the third measurement facet in a data matrix (in addition to the objects of measurement and one’s focal constructs) is not necessarily a measurement method (Lance et al., 2009). This is one key difference between our review and most previous reviews and one that we believe is at least partially responsible for the difference in esti- mated magnitudes of method effects across these reviews. These differences are especially evident when viewed in light of results from large-scale studies that have applied MTMM analyses to per- formance ratings and have routinely estimated substantial proportions of variance due to raters (Hoffman et al., in press; Scullen et al., 2000). Indeed, we excluded 80 studies because the alleged

‘‘methods’’ actually represented substantive measurement facets (e.g., raters, AC exercises, and measurement occasions). In the end, we included 18 studies that used different tests or different response formats as methods. How do we know that these represent alternative measurement meth- ods and not something substantively more meaningful? Actually, we do not, but at present have no reason to suspect that they do. Thus, this is our second call for future research on the fundamental nature of method effects. Some research (cited earlier) has suggested strongly that certain measure- ment facets that were once thought of as mere measurement methods are now better understood as representing substantive influences of focal constructs in their respective nomological networks. The prototype multidimensional measurement classification system by Lance et al. (2009) may prove useful in these efforts. On the other side of the coin, there are various response style phenomena (Greenleaf, 1992a) such as acquiescence (e.g., Bentler, Jackson, & Messick, 1971), differential use of positively versus negatively worded items (Weems, Onwuegbuzie, Schreiber & Eggers, 2003), central tendency (e.g., Weems & Onwuegbuzie, 2001), and extreme response style (e.g., Greenleaf,

1992b) that also deserve additional research as source of method variance independent of the mea- surement method per se.2

One limitation to this study is the relatively small number of MTMM studies that we eventually retained for analysis, just over 1% of the studies that we initially identified. Understandably, many citations to the D. T. Campbell and Fiske (1959) article did not actually conduct a MTMM study. However, many that did failed to report the MTMM matrix. In the same spirit of cumulating knowl- edge as in meta-analysis (Hunter & Schmidt, 2004), we urge future researchers to report the entire MTMM matrix, along with descriptive statistics (Ms and SDs) in their published (and unpublished) research. In addition, we found that a number of otherwise eligible MTMM matrices did not con- verge in a proper solution under a CTCM model CFA parameterization. As we acknowledged ear- lier, the CTCM does experience empirical underidentification problems (Brannick & Spector, 1990; Kenny & Kashy, 1992) but only under the CTCM model are trait and method factor loadings and correlations estimable; no other CFA model for MTMM data currently on the market estimates these quantities. Thus, we were faced with the unfortunate situation of being able to obtain unbiased para- meter estimates in the complete MTMM model for only a subset of the available MTMM matrices.

Conclusion

For nearly 50 years, organizational and behavioral researchers have bemoaned the influence of com- mon method effects as posing a serious threat to substantive conclusions. We reevaluated the effects of common method variance by reestimating method effects using what we have argued is a more appropriate sample of studies and by considering the counteracting effects of unreliability and com- mon method variance. In contrast to conventional wisdom, common method effects do not appear to be so large as to pose a serious threat to organizational research, especially when the counteracting effects of measurement error are considered.

Notes

1. For a CTCM CFA model to be identified, the MTMM matrix being analyzed must have at least three traits crossed with three methods, two traits crossed with four methods, or four traits crossed with two methods. The 16 studies that were eliminated on this basis were dimensioned smaller than the minimum required.

2. We thank an anonymous reviewer for alerting us to this point.

Declaration of Conflicting Interests

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

Financial Disclosure/Funding

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

References

Allen, J. G., & Hamsher, J. H. (1974). The development and validation of a test of emotional styles. Journal of

Consulting and Clinical Psychology, 42, 663-668.

Arora, R. (1982). Validation of an S-O-R model for situation, enduring and response components of involve- ment. Journal of Marketing Research, 19, 505-516.

Arthur, W. Jr., Woehr, D. J., & Maldegan, R. (2000). Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical reexamination of the assessment center construct related validity paradox. Journal of Management, 26, 813-835.

Avolio, B. J., Yammarino, F. J., & Bass, B. M. (1991). Identifying common methods variance with data col- lected from a single source: An unresolved sticky issue. Journal of Management, 17, 571-587.

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

Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. Psychological Bulletin, 76, 186-204.

Birnbaum, P. H., Farh, J., & Wong, G. Y. Y. (1986). The job characteristics model in Hong Kong. Journal of

Applied Psychology, 71, 598-605.

Borman, W. C. (1974). The rating of individuals in organizations; An alternative approach. Organizational

Behavior and Human Performance, 12, 105-124.

Bowler, M. C., & Woehr, D. J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. Journal of Applied Psychology, 91, 1114-1124.

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

Buckley, M. R., Cote, J. A., & Comstock, S. M. (1990). Measurement errors in the behavioral sciences: The case of personality/attitude research. Educational and Psychological Measurement, 50, 447-474.

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 outgoing editor. Journal of Applied Psychology, 67,

691-700.

Cattell, R. B. (1966). The data box: Its ordering of total resources in terms of possible relational systems. In R.

B. Cattell (ed.), Handbook of multivariate experimental psychology (pp. 67-128). Chicago: Rand McNally. Choi, N. (2004). A psychometric examination of the personal attributes questionnaire. Journal of Social

Psychology, 144, 348-352.

Cole, D. A. (2006). Coping with longitudinal data in research on developmental psychopathology. International

Journal of Behavioral Development, 30, 20-25.

Conway, J. M., & Huffcutt, A. I. (1997). Psychometric properties of multisource performance ratings: A meta- analysis of subordinate, supervisor, peer, and self-ratings. Human Performance, 10, 331-360.

Conway, J. M., Lievens, F., Scullen, S. E., & Lance, C. E. (2004). Bias in the correlated uniqueness model for

MTMM data. Structural Equation Modeling, 11, 535-559.

Cote, J. A., & Buckley, M. R. (1987). Estimating trait, method and error variance: Generalizing across 70 con- struct validation studies. Journal of Marketing Research, 26, 315-318.

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.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Fort Worth, TX: Harcourt, Brace, Jovanovich.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52,

281-302.

D’Mello, S. D., & Graesser, A. (2009). Automatic language detection of learner’s affect from gross body lan- guage. Applied Artificial Intelligence, 23, 123-150.

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. Structural Equation Modeling, 13, 440-464.

Doty, D. H., & Glick, W. H. (1998). Common methods bias: Does common methods variance really bias results? Organizational Research Methods, 1, 374-406.

Flamer, S. (1983). Assessment of the MTMM matrix validity of Likert scales via confirmatory factor analysis.

Multivariate Behavioral Research, 18, 275-308.

Franic, D. M., & Bothe, A. K. (2008). Psychometric evaluation of condition-specific instruments used to assess health-related quality of life, attitudes, and related constructs in stuttering. American Journal of Speech- Language Pathology, 17, 60-80.

Frese, M., & Zapf, D. (1988). Methodological issues in the study of work stress: Objective vs. subjective mea- surement of work stress and the question of longitudinal studies. In C. L. Cooper & R. Payne (Eds.), Causes, coping and consequences at work (pp. 375-410). Chichester: Wiley.

Greenleaf, E. A. (1992a). Improving rating scale measures by detecting and correcting bias components in some response styles. Journal of Marketing Research, 29, 176-188.

Greenleaf, E. A. (1992b). Measuring extreme response style. Public Opinion Quarterly, 56, 328-351. Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. Organiza-

tional Behavior and Human Performance, 16, 250-279.

Heisenberg, W. (1927). U¨ ber den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. [About the graphic content of quantum theoretic kinematics and mechanics]. Zeitschrift fu¨r Physik, 43,

172-198.

Hewitt, A. K., Foxcroft, D. R., & MacDonald, J. (2004). Multitrait-multimethod confirmatory factory analysis of the attributional style questionnaire. Personality and Individual Differences, 37, 1483-1491.

Hoffman, B. J., Lance, C. E., Bynum, B. H., & Gentry, W. A. (in press). Rater source effects are alive and well after all. Personnel Psychology.

Hoffman, B. J., & Woehr, D. J. (2009). Disentangling the meaning of multisource feedback: Expanding the nomological network surrounding source and dimension factors. Personnel Psychology, 62, 735-765.

Hunter, J. E., & Schmidt, F. L. (2004). Methods of meta-analysis: Correcting error and bias in research find- ings (2nd ed.). Thousand Oaks, CA: Sage.

Johnson, S. M., Smith, P. C., & Tucker, S. M. (1982). Response format of the Job Descriptive Index: Assess- ment of reliability and validity by the multitrait-multimethod matrix. Journal of Applied Psychology, 67,

500-505.

Jo¨ reskog, K. G., & So¨ rbom, D. (1993). LISREL 8 user’s reference guide. Chicago: Scientific Software. Keaveny, T. J., & McGann, A. F. (1975). A comparison of behavioral expectation scales and graph rating

scales. Journal of Applied Psychology, 60, 695-703.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. Psychological Bulletin, 112, 165-172.

Kenny, D. A., & Zautra, A. (2001). Trait-state models for longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), New methods for the analysis of change (pp. 243-263). Washington, DC: American Psychological Association.

King, L. M., Hunter, J. E., & Schmidt, F. L. (1980). Halo in a multidimensional forced-choice performance evaluation scale. Journal of Applied Psychology, 65, 507-516.

Kothandapani, V. (1971). Validation of feeling, belief, and intention to act as three components of attitude and their contribution to prediction of contraceptive behavior. Journal of Personality and Social Psychology, 19,

321-333.

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

Lance, C. E. (2008b). Where have we been, how did we get there, and where shall we go? Industrial and Orga- nizational Psychology, 1, 140-146.

Lance, C. E., Baranik, L. E., Lau, A. R., & Scharlau, E. A. (2009). If it ain’t trait it must be method: (Mis)ap- plication 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., Baxter, D., & Mahan, R. P. (2006). Multi-source performance measurement: A reconceptualiza- tion. In W. Bennett, C. E. Lance, & D. J. Woehr (Eds.), Performance measurement: Current perspectives and future challenges (pp. 49-76). Mahwah, NJ: Erlbaum.

Lance, C. E., Hoffman, B. J., Gentry, W. A., & Baranik, L. E. (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., Lambert, T. A., Gewin, A. G., Lievens, F., & Conway, J. M. (2004). Revised estimates of dimen- sion and exercise variance components in assessment center post-exercise dimension ratings. Journal of Applied Psychology, 89, 377-385.

Lance, C. E., Newbolt, W. H., Gatewood, R. D., Foster, M. R., French, N., & Smith, D. E. (2000). Assessment center exercise factors represent cross-situational specificity, not method bias. Human Performance, 13,

323-353.

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

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.

Lance, C. E., Teachout, M. S., & Donnelly, T. M. (1992). Specification of the criterion construct space: An application of hierarchical confirmatory factor analysis. Journal of Applied Psychology, 77, 437-452.

Lance, C. E., & Vandenberg, R. J. (2009). Statistical and methodological myths and urban legends: Doctrine, verity, and fable in organizational and social research. New York: Routledge.

Lievens, F., Chasteen, C. S., Day, E. A., & Christiansen, N. D. (2006). Large-scale investigation of the role of trait activation theory for understandning assessment center convergent and discriminant validity. Journal of Applied Psychology, 91, 247-258.

Lievens, F., & Conway, J. M. (2001). Dimension and exercise variance in assessment center scores: A large- scale evaluation of multitrait-multimethod studies. Journal of Applied Psychology, 86, 1202-1222.

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

London, M., & Tornow, W. W. (1998). Maximizing the value of 360-degree feedback. Greensboro, NC: Center for Creative Leadership.

Meehl, P. E. (1990a). Why summaries of research on psychological theories are often uninterpretable. Psycho- logical Reports, 66, 195-244.

Meehl, P. E. (1990b). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. Psychological Inquiry, 1, 108-141.

Messiah, A. (1961-1962). Quantum mechanics. New York: Interscience.

Mosher, D. L. (1966). The development and multitrait-multimethod matrix analysis of three measures of three aspects of guilt. Journal of Consulting Psychology, 30, 25-29.

Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R., & Hezlett, S. A. (1998). Trait, rater, and level effects in 360-degree performance ratings. Personnel Psychology, 51, 557-576.

Ostroff, C., Kinicki, A. J., & Clark, M. A. (2002). Substantive and operational issues of response bias across levels of analysis: An example of climate-satisfaction relationships. Journal of Applied Psychology, 87,

355-368.

Palmer, J. K., & Loveland, J. M. (2004). Further investigation of the psychometric properties of saucier’s big five ‘‘mini-marker:’’ Evidence for criterion and construct validity. Individual Differences Research, 2,

231-238.

Pitoniak, M. J., Sireci, S. G., & Leucht, R. M. (2002). A multitrait-multimethod validity investigation of scores from a professional licensure examination. Educational and Psychological Measurement, 62, 498-516. Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in beha-

vioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychol- ogy, 88, 879-903.

Quilty, L. C., Oakman, J. M., & Risko, E. (2006). Correlates of the Rosenberg Self-Esteem Scale method effects. Structural Equation Modeling, 13, 99-117.

Rindskopf, D. (1983). Parameterizing inequality constraints on unique variances in linear structural models.

Psychometrika, 48, 73-83.

Rose, G. L., & Andiappan, P. (1978). Sex effects on managerial hiring decisions. Academy of Management

Journal, 21, 104-112.

Sackett, P. R., & Dreher, G. F. (1982). Constructs and assessment center dimensions: Some troubling findings.

Journal of Applied Psychology, 67, 401-410.

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.

Scullen, S. E., Mount, M. K., & Goff, M. (2000). Understanding the latent structure of job performance ratings.

Journal of Applied Psychology, 85, 956-970.

Seligson, J. L., Huebner, E. S., & Valois, R. F. (2003). Preliminary validation of the brief multidimensional student life satisfaction scale (bmslss). Social Indication Research, 61, 121-145.

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. (2006). Method variance in organizational research: Truth or urban legend? Organizational

Research Methods, 9, 221-232.

Stevens, S. S. (1946). On the theory of scales of measurement. Science, 103, 677-680.

Thompson, V. C. & Berven, N. L. (2002). Development of an instrument to measure cultural attitudes and beha- viors affecting vocational rehabilitation. Rehabilitation Counseling Bulletin, 45, 76-104.

Vandenberg, R. J. (2006). Statistical and methodological myths and urban legends: Where, pray tell, did they get this idea? Organizational Research Methods, 9, 194-201.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job perfor- mance ratings. Journal of Applied Psychology, 81, 557-574.

Weems, G. H., & Onwuegbuzie, A. J. (2001). The impact of midpoint responses and reverse coding on survey data. Measurement and Evaluation in Counseling and Development, 34, 166-176.

Weems, G. H., Onwuegbuzie, A. J., Schreiber, J. B., & Eggers, S. J. (2003). Characteristics of respondents who respond differently to positively and negatively worded items on rating scales. Assessment & Evaluation in Higher Education, 28, 587-607.

Wherry, R. J., & Bartlett, C. J. (1982). The control of bias in ratings: A theory of rating. Personnel Psychology,

35, 521-551.

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 per-

ceptions at work: Reality or artifact? Journal of Applied Psychology, 74, 462-468.

Woehr, D. J., & Arthur, W. Jr. (2003). The construct related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. Journal of Management, 29, 231-258.

Yang, W., Lance, C. E., & Hui, H. C. (2006). Psychometric properties of the Chinese self-directed search (1994 ed.). Journal of Vocational Behavior, 68, 560-576.

Bios

Charles E. Lance is a Professor of I/O psychology at The University of Georgia and conducts research in the areas of performance measurement, assessment center validity, research methods, and structural equation mod- eling. Dr. Lance is a Fellow of the Society for Industrial and Organizational Psychology (SIOP) and the Amer- ican Psychological Association and is a member of the Society for Organizational Behavior. He is currently Associate Editor of ORM, and on the editorial board of Personnel Psychology, Human Performance, and Group

& Organization Management.

Bryan L. Dawson, MS, is currently a graduate student in the Applied Psychology doctoral program at The Uni- versity of Georgia. His research focuses on managerial decision making procedures regarding racialized sexual harassment, particularly how the role of victim race may affect perceived severity of harassment in the growing Latino workforce. In addition, Bryan is currently investigating the behavioral affects of positive and negative beliefs of Affirmative Action policies in selection, training and promotion environments.

David B. Birkelbach, MS, is currently a graduate student in the Applied Psychology doctoral program at The University of Georgia. His research focuses on methodological issues concerning 360 designs, assessment cen- ters, and developing novel methods to measure work performance, particularly to determine the feasibility of military personnel productivity and contribution to mission readiness indicators. In addition, David is exploring broad-based organizational analysis and interventions with an emphasis on work design and skill-transfer in merger/acquisitions.

Brian J. Hoffman is an assistant professor in the Applied Psychology at the University of Georgia. His research focuses on organizational leadership, performance evaluations, and methods of managerial assessment, includ- ing assessment centers and multisource performance ratings.