ON THE ANALYSIS OF MULTITRAIT-MULTIMETHOD
MATRICES IN CONSUMER RESEARCH

Working Paper #643

Richard P. Bagozzi
and
Youjae Yi
University of Michigan

ON THE ANALYSIS OF MULTITRAIT-MULTIMETHOD MATRICES
IN CONSUMER RESEARCH

ABSTRACT

This paper examines alternative procedures for analyzing multitrait-multimethod matrices: the Campbell-Fiske procedure, confirmatory factor analysis, and the direct product model. The implicit assumptions, as well as strengths and weaknesses, of each approach are presented and the implications discussed. It is proposed that one should carefully examine model assumptions, individual parameters, various diagnostic indicators, as well as overall model fits. The implications of these recommendations are illustrated through reanalyses of data from earlier studies of consumer behavior. Potentially misleading conclusions in these studies are corrected in demonstrating the procedures. The results show that methods often have multiplicative effects, supporting the direct product model which has not been used in consumer research before. The use of multiple-method, multiple-measure approaches is also advocated by highlighting the limitations of single-method, single-measure approaches in theory testing.

* Richard P. Bagozzi is the Dwight F. Benton Professor of Marketing and Behavioral Science in Management and Youjae Yi is Assistant Professor of Marketing, School of Business Administration, The University of Michigan, Ann Arbor, MI 48109. The authors thank the editor and three anonymous JCR reviewers for their helpful comments on previous versions of this article. The financial assistance of the Michigan Business School is gratefully acknowledged.

Any measure often reflects not only a construct of interest but also measurement error. Measurement error occurs commonly in practice and is recognized as a serious problem throughout the social sciences (e.g., Fiske 1982; Peter 1981). The problem of measurement error is especially serious in consumer research "because the majority of measures used are based upon constructs that are abstract and difficult to measure" (Cote and Buckley 1988, p. 579). Cote and Buckley (1987) have indeed shown that abstract constructs are more difficult to measure and more prone to measurement error than concrete constructs; for example, measurement error accounted for 70 percent of variance in attitude measures. Therefore, it is important to understand the implications of measurement error in consumer research.

Measurement error can be partitioned into random error and systematic error such as method variance. Method variance refers to variance attributable to the measurement method rather than to the construct of interest. The term "method," which refers to the form of the measurement, is used in a broad sense that encompasses potential influences at different levels of abstraction such as the content of specific items, scale type, response format, and the general context (Fiske 1982, pp. 81-84). For example, methods can be specific to an item (e.g., item wording), similar to specific or unique factors in classical test score theory but different from common method factors (e.g., scale types). At a more abstract level, method effects might be interpreted in terms of response biases such as halo effects, social desirability, acquiescence, leniency effects, or yea-nay saying.

Each of the two measurement error components can have serious confounding influences on empirical results and yield potentially misleading conclusions (e.g., Campbell and Fiske 1959). Cote and Buckley (1988) discuss such effects of measurement errors in the context of consumer research. Random error tends to attenuate the observed relationships among variables in statistical analyses. Consequently, errors in inference, especially Type II errors, are likely to occur in the presence of random error. In statistical analyses with simultaneous equations, random error can result in Type I as well as Type II errors (Bagozzi 1990).

Method variance may also bias results by inflating or deflating the correlations among constructs. More specifically, method effects inflate the relationship between two measures when the correlation between the methods is higher than the observed correlation between the measures with method effects removed. This fact suggests that the use of similar or highly correlated methods tends

to inflate the correlation between constructs. Especially when the same method is used to measure different constructs, shared method variance would always inflate the observed correlation because the correlation between the methods is 1.0 (which will be higher than any possible correlation between the measures). In contrast, method factors will attenuate the observed relationship when the correlation between the methods is lower than the observed correlation between the measures with method effects removed. However, the observed relationship will not be affected by methods when the correlation between the methods is the same as the observed correlation between the measures (see Cote and Buckley 1988 for more details). In sum, the observed relationship between measures is contaminated by measurement errors (i.e., random error and method factors) that are irrelevant to the construct but inevitable in most measurement situations.

Since contamination due to random error and method factors provides potential threats to the validity of research findings, it is important to validate measures and disentangle the distorting influences of these errors before testing theory. This can be achieved by using multiple measures and multiple methods in measurement (Campbell and Fiske 1959). The triangulation so produced permits a decomposition of information into that due to a theoretical phenomenon of interest, method effects, and random error. Using a single measure does not permit one to take measurement error into account in analyses. Likewise, with a single method one cannot distinguish trait (i.e., substantive) variance from unwanted method variance, because each attempt to measure a trait is contaminated by irrelevant aspects of the method employed.

Given multiple measures obtained with multiple methods, construct validity can be assessed with the multitrait-multimethod (MTMM) matrix, the correlation matrix for different concepts (traits) when each of the concepts is measured by different methods. That is, the MTMM matrix allows researchers to assess the extent of true relationship among traits in the presence of both method variance and random error. Typically construct validity is defined broadly as the extent to which an operationalization measures the concept that it is supposed to measure (Peter 1981). Campbell and Fiske (1959) proposed that two aspects of construct validity should be assessed in the analysis of MTMM data: convergent and discriminant validity. Convergent validity is the degree to which multiple attempts to measure the same concept are in agreement. Discriminant validity is the degree to which measures of different concepts are distinct. Without determining construct validity with multiple

measures and multiple methods, one cannot estimate and correct for the confounding effects of random error and method variance, and the results of theory testing may be ambiguous.

Despite the importance of construct validation, the MTMM design has been used infrequently in consumer research. Only a few studies could be found in the Journal of Consumer Research since its inception 17 years ago (e.g., Foxman, Tansuhaj, and Ekstrom 1989; Seymour and Lessne 1984). Also, very few related consumer research studies can be found in the Journal of Marketing Research over the 26 years of its life (e.g., Arora 1982; Menezes and Elbert 1979).

What accounts for the dearth of application of MTMM designs in consumer research? Undoubtedly one explanation is the difficulty in obtaining multiple measures of each concept in one's theory and using different methods to do so. A second reason might be that consumer researchers do not fully appreciate the consequences of failing to use multiple measures or multiple methods (i.e., attenuation due to unreliability or confounding due to method variance). Or they might believe that the amount of method and error variance is minimal in typical research settings so that these consequences will be negligible.

However, such a position is not justified according to empirical evidence. In the confirmatory factor analysis of 70 published data sets from various disciplines, Cote and Buckley (1987) found that measurement error is a serious problem in many research settings. On average, measures contained 41.7% trait variance, 26.3% method variance and 32% random error. That is, more than 50 percent of the variance in measures was due to measurement error. Furthermore, significant method variance was found for all but three of the studies examined. The results indicate that measurement errors are not only prevalent but also relatively large. Similarly, in the confirmatory factor analysis of 11 data sets from the organizational psychology literature, Williams, Cote and Buckley (1987) found that method variance accounted for approximately 25% of the variation in measures. These findings raise serious questions about the assumption of minimal measurement error.

Another reason for so little use of MTMM designs in consumer research, and the one primarily addressed in this paper, is the absence of clear standards for analyzing and interpreting MTMM data. The original criteria offered by Campbell and Fiske are both imprecise and overly restrictive. For example, they fail to provide a quantitative estimate of the degree to which the requirements are met, and assume that all measures have equal reliabilities. Other procedures such as confirmatory factor

analysis (CFA) and the direct product model (DPM) have not been systematically used or discussed in an integrative way in consumer research. The result is that researchers know little about how and when to apply these methodologies for analyzing MTMM data.

One purpose of the present paper is to provide consumer researchers with better insights as to how to model and estimate the confounding influences of measurement error in substantive research by introducing, comparing, and contrasting these procedures. For this purpose, we identify the implicit assumptions and restrictions associated with each procedure, examine the extent to which they can be met, explicate strengths and weaknesses, and discuss the implications for when one approach would be more appropriate than the others. A second is to illustrate these procedures so that researchers can understand how to apply them. In this regard, we demonstrate the procedures by reanalyzing data from consumer behavior studies and correct potentially misleading conclusions in these studies. Furthermore, we provide a detailed description of how to apply the methods (e.g., input specification for the DPM). Another purpose is to show that the analysis of MTMM data is a complex process involving a number of criteria and that numerous pitfalls need to be recognized. It is proposed that one should carefully examine model assumptions, individual parameters, various diagnostic indicators as well as overall model fits. Finally, we advocate the use of multi-method multi-measure approaches in consumer research to overcome the limitations of single-method single-measure approaches by highlighting the consequences of neglecting or not properly modeling measurement error in theory testing.

THE CAMPBELL AND FISKE PROCEDURE

Campbell and Fiske (1959) specify four criteria for evaluating any MTMM. First, convergent validity is achieved when the (monotrait-heteromethod) correlations between attempts to measure the same concept with different methods are "significantly different from zero and sufficiently large" (Campbell and Fiske 1959, p. 82). The next three criteria are necessary conditions for the establishment of discriminant validity. Namely, discriminant validity is achieved when a) the (monotrait-heteromethod) correlations between attempts to measure the same concept with different methods are greater than the (heterotrait-heteromethod) correlations between attempts to measure different concepts with different methods, b) the (monotrait-heteromethod) correlations between attempts to measure the same concept with different methods are greater than the (heterotrait-

monomethod) correlations between different concepts measured with common methods, and c) similar patterns of correlations result within each of the matrices formed by correlating measures of different concepts obtained by the same methods and by correlating measures of different concepts obtained by different methods. See Browne (1984) for a discussion on the implications of these criteria for modelling trait and method effects.

A number of problems can be identified with Campbell and Fiske's criteria (e.g., Peter 1981; Widaman 1985). First, no precise standards are provided for ascertaining when any particular criterion is met. Instead, only rules of thumb are offered which depend on a qualitative assessment of the number of confirming and disconfirming incidents in the MTMM. Second, it is not possible to assess the separate amounts of trait, method, and random error in the data. Rather these are all confounded by examining only the raw correlations. Third, the following assumptions are implicitly made: "there are no correlations between trait and method factors; all traits are equally influenced by method factors; and ...method factors are uncorrelated" (Schmitt and Stults 1986, p. 2). The latter two assumptions are frequently violated, and even the first may not hold under particular conditions discussed later in the paper. Each of these problems limits the usefulness of Campbell and Fiske's procedures for assessing construct validity. We turn now to a method that overcomes the first two problems and the last two parts of the third.

CONFIRMATORY FACTOR ANALYSIS

The general form for the CFA model applied to the MTMM can be expressed through two sets of equations (e.g., Werts, Jöreskog, and Linn 1972):

$$\underset{\sim}{y} = [\underset{\sim}{\Lambda_T} \, \underset{\sim}{\Lambda_M}] \begin{bmatrix} \underset{\sim}{\eta_T} \\ \underset{\sim}{\eta_M} \end{bmatrix} + \underset{\sim}{\epsilon} \tag{1}$$

$$\underset{\sim}{\Sigma} = \underset{\sim}{\Lambda_T} \underset{\sim}{\Psi_T} \underset{\sim}{\Lambda'_T} + \underset{\sim}{\Lambda_M} \underset{\sim}{\Psi_M} \underset{\sim}{\Lambda'_M} + \underset{\sim}{\theta} \tag{2}$$

where $\underset{\sim}{y}$ is a vector of r x s measures for r traits by s methods, Λ_T and Λ_M are factor loading matrices for traits and methods, respectively (defined below), η_T and η_M are vectors of r traits and s method factors, respectively, ϵ is a vector of residuals for $\underset{\sim}{y}$, Σ is the implied covariance matrix for $\underset{\sim}{y}$, Ψ_T and Ψ_M are intercorrelation matrices for traits and methods, respectively, θ is the vector of unique variances for ϵ, $\Lambda_T = [\Lambda_1, \Lambda_2, ..., \Lambda_r]'$, Λ_i is a diagonal matrix with factor loadings corresponding to the measures of the i-th trait, and

$$\Lambda_M = \begin{bmatrix} \lambda_1 & 0 & . & . & . & 0 \\ 0 & \lambda_2 & . & . & . & 0 \\ . & . & . & & & . \\ . & . & . & . & & . \\ . & . & & . & & 0 \\ 0 & 0 & . & . & 0 & \lambda_s \end{bmatrix}$$

where λ_j is a vector of factor loadings corresponding to the measures obtained by the j-th method.

Application of the CFA model to MTMM data permits one to test for, and partition variance into, trait, method, and random error. These reside, respectively, in the squared factor loadings for Λ_T and Λ_M and in θ.

Following Widaman (1985; see also Cote and Buckley 1987), four CFA models can be tested and compared to yield meaningful tests of hypotheses:

Model 1: the model hypothesizing that only unique variances are free (termed the null model). This model implies that variation in measures is explained only by random error (no trait or method factors).

Model 2: the model hypothesizing that variation in measures can be explained completely by traits plus random error (termed the trait-only model).

Model 3: the model hypothesizing that variation in measures can be explained completely by methods plus random error (termed the method-only model)

Model 4: the model hypothesizing that variation in measures can be explained completely by traits, methods, and random error (termed the trait-method model).

Model 4 is, in fact, the hypothesis implied by Equations 1 and 2. This model implies that both trait and method factors are needed to explain the variance in the measures. Models 1-3 are special cases formed by constraining certain parameters of Model 4. The null model is nested in both the method-only and trait-only models, whereas the method-only and trait-only models are nested in the trait-method model. Consequently, chi-square difference tests can be used to test whether trait, method, or trait and method variance are present. Specifically, a test of trait variance is provided by comparing chi-square tests between Models 1 and 2 and between Models 3 and 4. Similarly, a test of method variance is provided by comparing Models 1 and 3 as well as Models 2 and 4. Only the trait-only and trait-method models have a substantive interpretation. The null and method-only models are used simply for statistical comparison purposes in tests of hypotheses.

The advantages of the CFA approach over Campbell and Fiske's criteria are a) measures of the overall degree of fit are provided in any particular application (e.g., the chi-square goodness-of-fit test, adjusted goodness-of-fit index, root mean square residual, and standardized residuals), b) useful information is supplied as to if and how well convergent and discriminant validity are achieved (i.e., through chi-square difference tests for hierarchically nested models, the size of factor loadings for traits, and the estimates for trait correlations), and c) explicit results are available for partitioning variance into trait, method, and error (assuming that traits and methods are independent following the convention in previous research). Thus, the CFA approach provides more diagnostic information about reliability and validity, compared with Campbell and Fiske's criteria.

One key aspect of the CFA approach can be identified that potentially limits its usefulness. This limitation stems from the assumed structure of the model. The CFA model hypothesizes that variation in measures will be a linear combination of traits, methods, and error. That is, traits, methods and error are presumed to have additive effects on measures in the CFA model. This assumption may be a reasonable one such as when the effects of sharing a method do not vary with trait factors. In certain contexts, however, traits and methods may interact in the determination of measure variation. Let us examine the processes that can produce the interaction between methods and traits.

One process is called differential augmentation (Campbell and O'Connell 1967, 1982). According to this view, the multiplicative relation occurs such that "the higher the basic relationship between two traits, the more that relationship is increased when the same method is shared" (Campbell and O'Connell 1982, p. 95). A conventional position is that method factors add irrelevant systematic (method-specific, trait-irrelevant) variance to the observed relationships. That is, sharing a method is expected to increase the correlations between two measures above the true relationship (e.g., halo effects and response sets). However, not all relationships are exaggerated by sharing the method, but only relationships that are large enough to get noticed are likely to be exaggerated. Suppose, for example, that ratings are used as methods (Campbell and O'Connell 1967). Each rater might have an implicit theory about the co-occurrence of certain traits, which will lead to a rater-specific bias. In such cases, the stronger the "true" associations are between traits, the more likely they are to be noticed and exaggerated, thus producing the multiplicative method effect pattern.

Another possible process underlying the multiplicative method effects is differential attenuation (Campbell and O'Connell 1967, 1982). Campbell and O'Connell (1982) pointed out that "not sharing the same method dilutes or attenuates the true relationship, so that it appears to be less than it should be" (p. 95). That is, methods are seen as diluting trait relationships rather than adding irrelevant systematic variance. This view asserts that not sharing a method attenuates the observed correlations differently, depending on the level of true trait relationships. Campbell and O'Connell 1982 (pp. 100-106) provided an example of such effects where multiple occasions are used as methods. It is often found in longitudinal studies that correlations are lower for longer time lapses than shorter lapses, following a so-called autoregressive process. Accordingly, a high correlation between two traits will be more attenuated over time than a low correlation. In contrast, a correlation of zero can erode no further, and it remains zero when computed across methods (here occasions).

The above discussion suggests that methods may interact with traits in some circumstances. Campbell and O'Connell (1967) went so far as to suggest that such an interaction is "quite general in nature" (p. 421). If so, the CFA model will be inappropriate. An important question then arises: Which model should be used to investigate such interaction effects? In the next section, we will describe the direct product model (DPM) which is designed to represent the interaction between traits and methods (e.g., Browne 1984).[1]

THE DIRECT PRODUCT MODEL

Swain (1975) proposed the direct product model (DPM) to represent the multiplicative interaction of traits and methods in the MTMM:

$$\underset{\sim}{\Sigma} = \underset{\sim}{\Sigma}_M \otimes \underset{\sim}{\Sigma}_T \tag{3}$$

where $\underset{\sim}{\Sigma}$ is the covariance matrix of the observed variables, $\underset{\sim}{\Sigma}_M$ and $\underset{\sim}{\Sigma}_T$ are method and trait covariance matrices, respectively, and $\otimes$ indicates a right direct (Kronecker) product. This model expresses the covariance matrix of measurements as the direct product of a covariance matrix of methods and a covariance matrix of traits. However, it does not allow for measurement errors or different scales for different variables, which can limit the applicability of the model in many MTMM studies. Browne (1984, 1989) thus extended the DPM to incorporate unique variances and scale factors:

$$\underset{\sim}{\Sigma} = \underset{\sim}{Z} (\underset{\sim}{P}_M \otimes \underset{\sim}{P}_T + \underset{\sim}{E}^2) \underset{\sim}{Z} \tag{4}$$

where $\underset{\sim}{Z}$ is a nonnegative definite diagonal matrix of scale constants some of which are set equal to unity to achieve identification, $\underset{\sim}{P}_M$ and $\underset{\sim}{P}_T$ are the nonnegative definite matrices of method and trait correlations, whose elements are particular multiplicative components of common score correlations (i.e., correlations corrected for attenuation), and $\underset{\sim}{E}^2$ is a diagonal matrix indicating the amount of unique variances. Thus, this is a multiplicative model for true scores or common scores in the factor analysis sense. Browne (1985) has developed a program, MUTMUM, to estimate the parameters in Equation 4, compute standard errors, and provide chi-square goodness-of-fit tests. Under Equation 4 the correlation matrix corrected for attenuation will have a direct product structure:

$$\underset{\sim}{P}_C = \underset{\sim}{P}_M \otimes \underset{\sim}{P}_T \tag{5}$$

where $\underset{\sim}{P}_C$ is the disattenuated correlation matrix for the common scores. Note that this equation assumes a multiplicative structure for common scores, rather than for observed scores.

Campbell and Fiske's (1959) original criteria for convergent and discriminant validity have the following interpretations in the DPM (Browne 1984). Evidence for convergent validity is achieved when the correlations among methods in $\hat{\underset{\sim}{P}}_M$ are positive and large. The first criterion for discriminant validity is met when the correlations among traits in $\hat{\underset{\sim}{P}}_T$ are less than unity. The second criterion for discriminant validity is attained when the method correlations in $\hat{\underset{\sim}{P}}_M$ are greater than the trait correlations in $\hat{\underset{\sim}{P}}_T$. The final discriminant validity criterion is satisfied whenever the DPM holds. It can be shown that these interpretations result from the DPM specification.

Let $\rho(T_i M_k, T_j M_l)$ denote the correlation (corrected for attenuation) between the i-th trait measured with the k-th method and the j-th trait measured with the l-th method. From Equation 4 it can be seen that $\rho(T_i M_k, T_j M_l) = \rho(T_i, T_j) \rho(M_k, M_l)$. The criterion for convergent validity is that the monotrait-heteromethod correlation should be substantially greater than zero. When we look at the monotrait-heteromethod correlation, we see that $\rho(T_i M_k, T_i M_l) = \rho(M_k, M_l)$. Thus, convergent validity is achieved when method correlations are large and positive. The first criterion for discriminant validity is that the monotrait-heteromethod correlations, $\rho(T_i M_k, T_i M_l)$, should be greater than the corresponding heterotrait-heteromethod correlations, $\rho(T_i M_k, T_j M_l)$ and $\rho(T_j M_k, T_i M_l)$, for $i \neq j$. Because $\rho(T_i M_k, T_j M_l)/\rho(T_i M_k, T_i M_l) = \rho(T_j M_k, T_i M_l)/\rho(T_i M_k, T_i M_l) = \rho(T_i, T_j)$, this criterion is met when trait correlations are less than unity. The second criterion for discriminant validity is that the monotrait-heteromethod correlations, $\rho(T_i M_k, T_i M_l)$, should be higher than the

corresponding heterotrait-monomethod correlations, $\rho(T_i M_k, T_j M_k)$ and $\rho(T_i M_l, T_j M_l)$. Since $\rho(T_i M_k, T_j M_k)/\rho(T_i M_k, T_i M_l) = \rho(T_i M_l, T_j M_l)/\rho(T_i M_k, T_i M_l) = \rho(T_i, T_j) / \rho(M_k, M_l)$, this criterion is met when the method correlations are greater than the trait correlations. The final criterion is that all matrices of intertrait correlations should have the same pattern whichever methods are used. This criterion is met whenever the DPM holds, since the ratio $\rho(T_i M_k, T_j M_l)/ \rho(T_m M_k, T_n M_l) = \rho(T_i, T_j)/\rho(T_m, T_n)$ has the same value for any M_k and M_l.

The CFA model and DPM hypothesize different functional forms for trait and method effects: the former additive, the latter multiplicative. Thus, in principle, the models constitute alternative explanations for MTMM data. If one employs an additive (multiplicative) model when in fact a multiplicative (additive) model is correct, the model will be misspecified. The specification error will yields biased estimates for method and trait effects, and induce misleading conclusions about the construct validity of the measures. It is therefore important to decide which specification should be used. Although Campbell and O'Connell (1967, 1982) imply that trait and method interactions are the rule rather than the exception, it might be useful in any specific case to test both additive and multiplicative models to discover which process is at work.

METHOD

Data

Four data sets were selected to illustrate the range of results, problems, and anomalies typical of any determination of construct validity through the analysis of MTMMs. They were gathered from published sources of consumer behavior studies in the literature. Brief descriptions of the data sets follow. Our descriptions of the studies hereafter present the results that the authors found. Later, in our reanalysis, we point out the shortcomings in their analyses and/or interpretations as we go along.

Arora (1982). Subjects (96 undergraduate business students) were asked to express their attitudes toward their university on three methods: semantic differential, Likert, and stapel scales. The three traits were situational, enduring, and response involvement. Arora (1982) performed a CFA and found that three oblique traits and three orthogonal methods satisfactorily explained the MTMM data. Trait variance was generally high, whereas method and error variance were relatively low.

Foxman et al. (1989). One hundred and sixty-one family triads consisting of a father, mother, and adolescent child were asked to rate the general influence on purchasing by the child. The father,

mother, and child were treated as methods. The child's general influence was assessed in each of 7 broad categories treated as traits: suggesting a price range, going shopping with parents when looking for a product for family use, suggesting products, paying attention to new products, and learning the best buy. In our study, the first three traits are used for illustrative purposes. Foxman et al. performed a partial application of Campbell and Fiske's (1959) original criteria to their MTMM data. They claimed mixed support for convergent validity, although the magnitude of validity diagonal correlations were low (range: .04 to .35). No attempt was made to address discriminant validity.

Menezes and Elbert (1979). Likert, semantic differential, and stapel scales were used to rate retail store image on appearance, products, and prices. A fourth trait, service, was not included in our analyses because of its low reliability scores on the three methods (i.e., the reliabilities for service items were found to be .57, .54, and .53, respectively). Subjects were 250 evening business school students. Although Menezes and Elbert (1979) performed a CFA on their data, they did not present findings permitting one to assess convergent or discriminant validity. Further, because they allowed all traits and methods to freely intercorrelate, their model is not identified. As a consequence, interpretations of their results, even if they had been presented, would be misleading (see Widaman 1985, p. 7).

Seymour and Lessne (1984). In an effort to develop a spousal conflict arousal scale, Seymour and Lessne asked 90 "married persons" to respond to 20 issues, each measured by three methods: Likert items, mixed standard scales, and graphic rating scales. Four different traits were formed as suggested by an earlier factor analysis of Likert items answered by a different sample of respondents. The traits were interpersonal need, utility, involvement, and power. Seymour and Lessne (1984) applied a portion of Campbell and Fiske's (1959) classic criteria for assessing MTMMs and concluded that, although convergent validity was in evidence for all four traits, utility and involvement lacked discriminant validity because their average correlation was .78. In our reanalyses, we will focus on involvement, power, and interpersonal need for illustrative purposes.

Analytical Procedures

The LISREL 7 program was used to perform all analyses because of its wide availability and previous use by consumer researchers (Jöreskog and Sörbom 1988). The application of LISREL for

the CFA is straightforward and will not be elaborated on herein. The use of LISREL to operationalize the DPM is not straightforward and therefore will be briefly described.

Wothke and Browne (1990) show that the DPM can be reformulated as a linear model and programmed within the LISREL context. Specifically, Equation 4 can be written as a second-order confirmatory factor model as follows:

$$\Sigma = \Lambda \Gamma \phi \Gamma' \Lambda' \tag{6}$$

where $\Lambda = Z$, Γ is the partitioned matrix

$$\Gamma = (C_M \otimes I_t \mid I_{mt}) = (\Gamma_1 \mid \Gamma_2) \tag{7}$$

C_M is a square, lower triangular matrix chosen such that $P_M = C_M \, C'_M$, I_t and I_{mt} are identity matrices, and

$$\phi = \left(\begin{array}{c|c} I_M \otimes P_t & 0 \\ \hline 0 & E^2 \end{array} \right) = \left(\begin{array}{cc} \phi_1 & 0 \\ 0 & \phi_2 \end{array} \right) \tag{8}$$

The DPM can be easily restricted to suitable submodels. One version of the model is defined by the additional restriction

$$E^2 = E_M \otimes E_T \tag{9}$$

with E_M and E_T diagonal. By using the fact that any symmetric, nonnegative definite matrix can be expressed as the product of a square matrix and its transpose (e.g., Searle 1982), this restriction can be rewritten as follows:

$$E^2 = (E_M^{1/2} \otimes I_t)(I_m \otimes E_T)(E_M^{1/2} \otimes I_t)' = \Gamma_2 \, \phi_2 \, \Gamma_2' \tag{10}$$

Wothke and Browne (1990) discuss identification and estimation and provide an illustration. For identification of scale factor estimates, one equality constraint per method is required. For instance, one may select a trait and set all its scale factors in Z equal to unity. Alternatively, all diagonal elements of C_M can be constrained to unity. The two types of restriction may be suitably combined. Another restriction is required in order to fix the scale of the error components, since (a E_M) $\otimes$ (b E_T) = $E_M \otimes E_T$ for any a = 1/b. This may be achieved by constraining one element in either E_M or E_T to unity. P_T is directly estimated in the model, and standard errors of its elements will be available from the LISREL solution. However, the estimate of P_M is obtained by rescaling $\hat{C}_M \, \hat{C}'_M$ into a correlation matrix, and standard errors are not available from the LISREL output.

Because the causal diagram for the LISREL operationalization is quite cumbersome (e.g., there are 27 latent variables for the smallest model with three traits and three methods), we have not provided a figure. However, Appendix I contains a sample input specification for the DPM analysis on LISREL 7.

RESULTS

To provide as thorough and informative a presentation of findings as possible, we have chosen to discuss the results for each approach in sequence. The full extent of the ambiguities and trade-offs between the approaches will not become apparent until we compare the findings across approaches at the end of this section and in the Discussion.

Confirmatory Factor Analyses

Table 1 presents the results of the four models discussed earlier for examining trait and method effects and the corresponding nested tests.[2] The first thing to notice is that the CFA model with traits and methods fits all data sets well. This is shown in the Trait-Method cells in Table 1. Therefore in 4 of 4 data sets, the CFA model explains the MTMM data quite well. We stress that this conclusion is based on an interpretation of the chi-square goodness-of-fit tests alone. Later we will scrutinize additional goodness-of-fit measures and other diagnostic criteria.

Notice next that the introduction of either trait or method factors significantly drops the chi-square value in each data set, indicating that meaningful improvements over the null model are achieved (see chi-square difference tests at ends of first row and first column for each data set presented in Table 1). Also, the inclusion of both trait and method effects significantly improves the fits of models over the trait-only and method-only models, pointing to a need for both trait and method factors (see chi-square difference tests at ends of second row and second column in each data set). Although the magnitude of the chi-square difference tests suggests that traits explain more variance than methods in all cases, except for the data of Foxman et al. (1989), both traits and methods are needed in the final analyses.

In his critique of the CFA approach to MTMM analyses, Browne (1984, p. 5) asserts that "(n)o information is provided as to whether or not the Campbell-Fiske requirements are met." Although we disagree with this assessment with respect to discriminant validity, as discussed shortly, we concur that the CFA approach gives ambiguous information with regard to convergent validity. Nevertheless,

some information is provided that goes beyond the Campbell and Fiske criterion for convergent validity.

We feel that the decomposition of variance into trait, method, and error supplies useful information for assessing the degree of convergent validity. The extent of trait variation reflects the magnitude of shared variation for two or more measures of a common trait factor. Within the context of CFA, this variation has method and error variance removed from it. Because convergent validity is defined as the agreement among measures of the same trait assessed by different methods, variation uniquely explained by traits yields a quantitative indicator of the degree of convergent validity (Marsh and Hocevar 1988; Widaman 1985).

The ambiguity arises when one must decide how much variance is sufficient for the attainment of convergent validity. We propose that two levels of construct validity can be assessed. Weak convergent validity results when the factor loading on a measure of interest is statistically significant. Strong convergent validity is achieved when at least half of the total variation is trait variance (i.e., $\lambda^2 \geq .5$). Note that these assessments are made against the backdrop of the general evaluation of overall model fit. A satisfactory goodness-of-fit measure(s) implies that the hypothesis of underlying trait and method factors explains the data except for random fluctuations. Each of these rules is admittedly an arbitrary heuristic. Yet each goes further than Campbell and Fiske's (1959) criterion for convergent validity in that a quantitative measure is provided that partials out method and error variance and an overall goodness-of-fit measure is supplied. According to these rules of thumb, convergence (i.e., significant trait variance) is achieved for all measures in Arora (1982) and Seymour and Lessne (1984), but for only six of nine measures in Foxman et al. (1989) and seven of nine measures in Menezes and Elbert (1979).

The amount of variance due to traits ranges from .36 to .80, .23 to .92, .00 to 44, and .10 to .60, for the data of Arora (1982), Seymour and Lessne (1984), Foxman et al. (1989), and Menezes and Elbert (1979), respectively. Error variance ranges from .00 to .38, .00 to .25, .13 to .68, and .03 to .30 for these respective data sets. In terms of method variance, five of nine, four of nine, eight of nine, and nine of nine method factor loadings are significant in the data of Arora (1982), Seymour and Lessne (1984), Foxman et al. (1989), and Menezes and Elbert (1979), respectively. The amount of variance due to methods ranges from .01 to .30, .00 to .25, .01 to .79, and .16 to .74 for these

respective data sets. The full results for the partitioning of trait, method, and error variance for all measures are available on request from the authors.

We assessed discriminant validity by examining correlations among traits (Ψ_T) and methods (Ψ_M). A perfect correlation between traits would indicate that the traits are not discriminable. Discriminant validity among traits is achieved when the trait correlation differs significantly from 1.00 or when the chi-square difference test indicates that the two traits are not perfectly correlated (e.g., Schmitt and Stults 1986; Widaman 1985). Discriminant validity is attained for all the traits in Arora (1982) as well as Seymour and Lessne (1984). In contrast, all three traits in Foxman et al. (1989) lack discriminant validity. The first and second traits in Menezes and Elbert (1979) fall short of discriminant validity, but the third trait is distinct from the first two. The findings for the discrimination among methods show that the methods are distinct in Arora (1982) and Foxman et al. (1989). The confidence intervals for the three methods in Menezes and Elbert (1979) nearly reach 1.00. The first and third methods in Seymour and Lessne (1984) are not distinct.

In sum, a fairly detailed and seemingly clear picture of construct validity is provided for each data set by the CFA approach. However, we believe that the standard analysis--which relies upon the chi-square goodness-of-fit test, chi-share difference tests, a partitioning of variation into trait, method, and error, and an examination of trait correlations--can be misleading. To see this better, consider Table 2 which summarizes a number of additional diagnostics for each data set. The chi-square test, adjusted goodness-of-fit index (AGFI), and root mean squared residual (RMR) are overall measures of fit in the sense of expressing the discrepancy between the covariance matrix implied by the hypothesized model and the observed covariance matrix. Although these measures seem to suggest satisfactory solutions (in fact, all three indices suggest that each of the four data sets can be explained satisfactorily by the CFA specification), it is still possible that they might be contaminated by methodological artifacts. For example, the chi-square test could point to a satisfactory fit only because the test lacks statistical power. Likewise, when many trait and method factors are introduced into a model, as must be done to implement a structural equation model of the MTMM, the danger exists that satisfactory chi-square, AGFI, and RMR values may arise as a result of over-fitting.

[Tables 1 & 2 about here]

For these reasons, we recommend that two additional procedures be followed in any evaluation of MTMM results. First, the standardized residuals should be examined in any analysis. The standardized residuals are formed by taking the residuals from the observed and implied covariance matrices and dividing these residuals by their asymptotic standard errors. "Each standardized residual can be interpreted as a standard normal deviate and considered 'large' if it exceeds the value 2.58 in absolute value" (Jöreskog and Sörbom 1988, p. 32). One or more large standardized residuals indicate that a significant amount of variance remains unexplained and that the model may be misspecified. Standardized residuals can be printed out as an option on LISREL 7.

A second procedure that we advocate is an examination of each parameter for the presence of either improper or incongruous solutions. An improper solution is one that is either illogical or outside the range of conventional acceptability. Negative error variances, correlations greater than 1.00, and standardized factor loadings greater than 1.00 are examples. An incongruous solution is a parameter estimate that is highly unlikely, contradicts what would be expected on the basis of theoretical or methodological reasoning, or is in some way inexplicable. For instance, the presence of nonsignificant error terms often suggests over-fitting or misspecification biases because one normally anticipates at least a small amount of residual variance in social science data (e.g., Maxwell 1977, p. 58). Similarly, factor loadings for method effects and correlations among methods are sometimes inconsistent in the sense of yielding both positive and negative loadings on the same factor or producing contradictory associations among factors. These findings are typically uninterpretable. Browne (1984, p. 7) terms these "wastebasket parameters" to indicate that they are introduced to achieve a satisfactory goodness-of-fit but do not have a substantive interpretation (see also Kenny 1979, p. 154).

Applying the above criteria to our analyses, we obtain the summary of results shown in Table 2. Notice first that the Arora (1982), Menezes and Elbert (1979), and Seymour and Lessne (1984) analyses reveal 5, 1, and 1 large standardized residuals, respectively. An improper solution was found for the correlation between the first and second traits in Foxman et al. (1989), and the Arora (1982), Foxman et al. (1989), Menezes and Elbert (1979), and Seymour and Lessne (1984) solutions yield 3, 1, 1, and 3 nonsignificant estimates of θ_i, respectively.[3] Two additional incongruous (i.e., "wastebasket") parameter estimates were discovered for method factors in the Seymour and Lessne (1984) analyses.

Based on these findings with respect to the standardized residuals, improper solutions, and wastebasket parameters, we have reason to reject the CFA solutions for all the data sets. Thus, we reject the hypothesis of linear, additive effects for trait, method, and error for these data sets.

Direct Product Model Analyses

Table 3 presents a summary of the findings for the DPM applied to each data set. On the basis of the standard goodness-of-fit indicators, the DPM appears to fit the data of Foxman et al. (1989) and Menezes and Elbert (1979). The DPM fits poorly to the data of Arora (1982) and Seymour and Lessne (1984), as shown by the high chi-square values and RMR estimates and the relatively low AGFIs. An inspection of the standardized residuals reveals no large value in all the DPM analyses. Two error messages arose in the analyses of the Arora (1982) and Seymour and Lessne (1984) data, suggesting that two parameters were unidentified. Because the parameters in question were in fact theoretically identified, it is likely that the messages refer to empirical underidentification (Kenny 1979; Rindskopf 1984). Notice that these two data sets showed rather poor goodness-of-fit indicators. In addition, because the sample sizes are relatively small for these data ($\underline{n}$ = 96 and $\underline{n}$ = 90, respectively), it is likely that the chi-square tests validly point to meaningful discrepancies between the hypothesized DPMs and the data. Therefore, when all the goodness-of-fit indicators and diagnostics are taken into account, the evidence supports the DPM for Foxman et al. (1989) and Menezes and Elbert (1979).

With this as background, let us examine the specific criteria for assessing convergent and discriminant validity in the DPM. These are displayed in Table 4. Because the method correlations are relatively large, convergent validity is achieved for Arora (1982), Menezes and Elbert (1979), and Seymour and Lessne (1984). With two of three low method correlations and one moderate correlation in Foxman et al. (1989), we reject convergent validity therein.

[Tables 3 & 4 about here]

All three criteria for discriminant validity are met in Arora (1982), Menezes and Elbert (1979), and Seymour and Lessne (1984). The second criterion fails for Foxman et al. (1989). Note that standard errors of parameter estimates are available for ϕ_T but not ϕ_M, given that the correlations are formed as products of coefficients.[4] The three method correlations in Menezes and Elbert (1979) and one in Seymour and Lessne (1984) are rather large, implying that the corresponding methods are similar.

In sum, the DPM model is consistent with the data of Foxman et al. (1989) and Menezes and Elbert (1979). In the latter case, convergent and discriminant validity are established. The poor fits and the accompanying small sample sizes for the data in Arora (1982) and Seymour and Lessne (1984) make interpretations of convergent and discriminant validity problematic therein. In the next section we will attempt to reconcile the discrepancies between the CFA and DPM analyses and point out complementarities as well.

DISCUSSION

We have examined the nature of method effects (i.e., additive or multiplicative) for the four data sets by comparing two alternative models: CFA and DPM. To bring the presentation into perspective, we summarize the conclusions implied by the CFA and DPM analyses in Table 5. The conclusions in the table are based on a full interpretation of goodness-of-fit measures, parameter estimates, and the other diagnostics mentioned earlier. Let us first examine the conclusions about model fit. The DPM, but not the CFA model, fits the data of Foxman et al. (1989) and Menezes and Elbert (1979). Hence, the results tend to support the premise that variation in MTMM data can be explained by either additive or multiplicative method effects.

[Table 5 about here]

It can also be noted that both CFA and DPM were rejected for Arora (1982) and Seymour and Lessne (1984). Hence, any conclusions drawn here with respect to convergent and discriminant validity must be qualified with the foreknowledge that the models are misspecified. Although we have tested the possibility that methods and traits combine in an additive or multiplicative way, these results suggest that they might combine in still other ways. For instance, method and trait factors may interact inversely; that is, the higher the trait correlations are, the lower the method effects are (Campbell and O'Connell 1982). In the future, it will be useful to develop models that can represent such alternative processes.

If we were to have based our conclusions about model fit solely on the chi-square goodness-of-fit tests (see Tables 2 and 3), the CFA model would be accepted for 4 of 4 data sets, whereas the DPM would be accepted for 2 of 4 data sets. The CFA and DPM are at odds with respect to the data of Arora (1982) and Seymour and Lessne (1984) in that the former points to acceptance and the latter rejection. For the remaining two data sets, both the CFA and DPM fit well when we scrutinize only the chi-

square tests. Aside from overlooking the anomalies noted in the Results section, the use of only the chi-square goodness-of-fit test thus leads to ambiguous and contradictory results for many data sets. We again see the need for a deeper evaluation of individual parameter estimates, standardized residuals, and the additional diagnostics noted earlier.

If we examine the evidence for convergent and discriminant validity, we see that the CFA and DPM conclusions are generally in agreement (see Table 5). The one exception occurs for the analyses of the data in Menezes and Elbert (1979). Here the CFA points to mixed achievement of convergent and discriminant validity, whereas the DPM leads to a conclusion of satisfactory convergent and discriminant validity. Which conclusion should be accepted? Recall that the DPM gave a satisfactory fit to the data, whereas the CFA was rejected. From this perspective, more credence should be given to the conclusion from the DPM analysis. This result suggests that one should investigate the nature of method effects (i.e., additive or multiplicative) before assessing construct validity. Assessment of construct validity via an inappropriate model can produce misleading conclusions. For instance, if methods do interact with traits, as in the data of Menezes and Elbert (1979), the use of CFA models (which are misspecified) will yield biased estimates and misleading conclusions.

The choice between the CFA and DPM depends on the way that method and trait factors influence the variation in measures. The CFA model is appropriate when the effects of methods and traits are additive. In contrast, the DPM is appropriate when one has substantive reason for expecting interactive effects of methods and traits. For example, Jackson (1969) cites the case where acquiescence (a method) has a stronger effect when knowledge (a trait) is low than high. To date, most MTMM analyses have assumed the additive effects of methods and traits. As Campbell and O'Connell (1969, p. 424) argued, additive models of factor analysis have been untested by using factor analysis as the criterion to which the data should fit. However, the analysis procedures should be chosen based upon the fit of the models (e.g., CFA vs. DPM) to the data. Our analyses revealed that the multiplicative effects are plausible at least for 2 of 4 data sets examined, suggesting that more attention should be given to the DPM in MTMM analyses (cf. Lastovicka, Murry, and Joachimsthaler 1990).

We should mention one limitation of the DPM that may or may not be an issue, depending on the particular data at hand. The DPM criterion for convergent validity requires that the method correlations be substantial. Notice that this is a composite indicator which implicitly takes into account

the convergence among multiple measures for each trait. The CFA criterion for convergent validity is based on the amount of trait variance for each measure. When all measures tend to converge to the same degree, the CFA and DPM findings should agree. But when some measures converge to a significantly greater degree than others for a common trait, the CFA procedure retains this information better than the DPM. The suggested CFA criterion based on the trait factor loading provides diagnostic information as to which of the measures achieve convergent validity. Such diagnostic results can aid in item selection and point to candidates for inclusion and exclusion in future research. It thus appears that the DPM is less informative than the CFA with respect to convergent validity.

Some caveats of this study are in order. First, the multiple methods used in most data sets under investigation were fairly similar. Except for the Foxman et al. (1989) study, all methods were paper and pencil self-report forms filled out by the same respondents. In fact, the method correlations in Table 4 suggest that the methods were quite similar in all studies except Foxman et al. (1989). The results should be interpreted in light of this caveat. On the one hand, the most stringent tests of convergent validity require maximally dissimilar methods (Campbell and Fiske 1959). On the other hand, the most stringent tests of discriminant validity require maximally similar methods. Ideally, then, a construct validity study should employ both maximally similar and dissimilar methods. This, of course, multiplies the number of measures required.

The findings of this investigation might be limited in their generalizability, because they are based on four empirical data sets. The critical issue pertains to using the proper specification for trait and method effects. If trait and method factors interact, the use of CFA yields a misspecified model. The consequences and implications of such misspecifications would be better understood with simulations. It should also be mentioned that the rules suggested for evaluating convergent and discriminant validity are merely heuristics, and that these criteria will be affected by the model fit and the presence of wastebasket parameters. Further, it should be acknowledged that we faced difficulties in estimating the MTMM models: we had to choose starting values judiciously in order to obtain convergent solutions. This estimation problem, often experienced by researchers in this area (Schmitt and Stults 1986), might be serious when the MTMM design becomes complex.

Finally, we wish to stress that the conclusions drawn in this paper are based on commonly accepted statistical criteria. An important issue to consider is the practical relevance of the findings.

For each of the data sets, we computed the normed fit index for the trait-only CFA model compared with the null model (e.g., Bentler and Bonett 1980; Mulaik et al. 1989). The index gives the proportion of total information accounted for by the trait-only model from a practical standpoint. For our computations of the normed fit indices, we subtracted the appropriate degrees of freedom from their respective chi-square values to yield noncentralized estimates and thus remove the bias in small samples of the ordinary index (e.g., McDonald and Marsh 1990).

The noncentralized normed fit indices (NCNFIs) are .84, .63, .97, and .92, respectively, for the trait-only models of the data of Arora (1982), Foxman et al. (1989), Menezes and Elbert (1979), and Seymour and Lessne (1984). It has been suggested that values of the index greater than or equal to .90 indicate that the hypothesized model accounts for a sufficient amount of information from a practical perspective (e.g., Bentler and Bonett 1980). Thus, on this basis, the trait-only model is not sufficient for the data of Arora and Foxman et al. but is satisfactory for the data of Menezes and Elbert, and Seymour and Lessne. The introduction of method factors makes sense for the first two data sets but could be considered overfitting for the latter two. We stress that this conclusion is based on the interpretation of the NCNFI as a measure of practical relevance. Our analyses based on statistical criteria described above indicate that the trait-only model must be rejected for all data sets.

As a final interpretation of the four data sets previously analyzed in the consumer behavior literature, we note that our analyses result in markedly different conclusions from those stated by the original authors. Two of the data sets (Arora 1982; Seymour and Lessne 1984) showed poor model fits by both the CFA and DPM approaches. Thus, contrary to conclusions made in the original studies, no firm basis exists for interpreting convergent and discriminant validity. In the third data set (Foxman et al. 1989), the CFA model was rejected and the DPM described the data satisfactorily. However, both convergent and discriminant validity must be rejected based on the findings, and thus the substantive conclusions made in the original study rest on questionable assumptions. Finally, for the fourth study (Menezes and Elbert 1979), our analyses show that the procedure originally used was both performed and interpreted improperly. The proper conclusion is that the CFA model must be rejected. Significantly, we found that the DPM adequately describes the data, and that convergent and discriminant validity are achieved.

CONCLUSIONS

We suggest that single-method, single-measure approaches be replaced with mutiple-method, multiple-measure approaches in consumer research. Use of a single measure requires an assumption that variables are perfectly measured without error, which is quite unlikely in most situations. Failure to model explicitly such random measurement errors, where they are present, will attenuate the relationships among the variables in any test of a theory. Similarly, use of a single method does not allow for a rigorous assessment of construct validity, because trait variation is confounded with variation due to methods. Failure to model such systematic errors in measurement, where they are present, can lead to biased and inconsistent estimates of the key parameters. Therefore, consumer researchers should try harder to obtain multiple measures of constructs with multiple methods and model random error as well as systematic error before testing substantive hypotheses. Although such an attempt might be costly and time-consuming, it is necessary for disentangling the confounding influences of method variance and random error on research findings.

Given multiple measures and multiple methods, one can assess the validity and reliability of measurement by analyzing MTMM data. We have considered a procedure infrequently used and introduced a new procedure for analyzing MTMM data in consumer research (i.e., CFA and DPM). We have considered several assumptions that are made implicitly or explicitly under each approach. In analyzing MTMM data a researcher should explicitly consider the nature of the assumptions underlying each procedure, examine the plausibility of these assumptions, choose the most appropriate one, and communicate these assumptions clearly to the readers. We have found that method effects are sometimes multiplicative rather than additive, so that the usual CFA is inappropriate. This suggests that one should consider and test alternative models against data, rather than merely assuming that a model (e.g., CFA) is appropriate. We have also tried to show that the interpretation of findings from such procedures is a complex process involving a number of criteria. The dearth of work in the area--both in consumer research and psychometrics--precludes us from making broad generalizations. Yet, we feel that our presentation at least presents the major alternatives and suggests interpretive guidelines for future research.

Appendix I

The following is a sample LISREL 7 program specification of the direct product model for the MTMM data with three traits and three methods.

```
Title Direct Product Model
DA NI=9 NO=161 MA=KM
[Data]
MO NY=9 NE=9 NK=18 LY=DI, FR GA=FU, FR PH=SY, FR PS=ZE BE=ZE TE=ZE
EQ LY 1 LY 4 LY 7
ST .8 LY 1 LY 4 LY 7
ST .7 LY 2 LY 3 LY 5 LY 6 LY 8
ST .9 LY 9
PA GA
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0
0 1 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0
0 0 1 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0
1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 1 0 0
0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 1 0
0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 1
VA 1 GA 1 1 GA 2 2 GA 3 3
EQ GA 4 1 GA 5 2 GA 6 3
EQ GA 4 4 GA 5 5 GA 6 6
EQ GA 7 1 GA 8 2 GA 9 3
EQ GA 7 4 GA 8 5 GA 9 6
EQ GA 7 7 GA 8 8 GA 9 9
VA 1 GA 1 10 GA 2 11 GA 3 12
EQ GA 4 13 GA 5 14 GA 6 15
EQ GA 7 16 GA 8 17 GA 9 18
ST .9 GA 4 1 GA 5 2 GA 6 3
ST .8 GA 4 4 GA 5 5 GA 6 6
ST .7 GA 7 1 GA 8 2 GA 9 3
ST 1 GA 7 4 GA 8 5 GA 9 6
ST .6 GA 7 7 GA 8 8 GA 9 9
ST .5 GA 4 13 GA 5 14 GA 6 15
ST .8 GA 7 16 GA 8 17 GA 9 18
```

PA PH

*

0

1 0

1 1 0

0 0 0 0

0 0 0 1 0

0 0 0 1 1 0

0 0 0 0 0 0 0

0 0 0 0 0 0 1 0

0 0 0 0 0 0 1 1 0

0 0 0 0 0 0 0 0 0 1

0 0 0 0 0 0 0 0 0 0 1

0 0 0 0 0 0 0 0 0 0 0 1

0 0 0 0 0 0 0 0 0 0 0 0 1

0 0 0 0 0 0 0 0 0 0 0 0 0 1

0 0 0 0 0 0 0 0 0 0 0 0 0 0 1

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1

EQ PH 2 1 PH 5 4 PH 8 7

EQ PH 3 1 PH 6 4 PH 9 7

EQ PH 3 2 PH 6 5 PH 9 8

EQ PH 10 10 PH 13 13 PH 16 16

EQ PH 11 11 PH 14 14 PH 17 17

EQ PH 12 12 PH 15 15 PH 18 18

VA 1 PH 1 1 PH 2 2 PH 3 3 PH 4 4 PH 5 5 PH 6 6 PH 7 7 PH 8 8 PH 9 9

ST .45 PH 2 1

ST .30 PH 3 1

ST .60 PH 3 2 PH 10 10

ST .50 PH 11 11 PH 12 12

OU RS NS SS TV SE

FOOTNOTES

1   Another limitation of the CFA model is that it confounds specific and error variance in measures. The error term in the CFA model is a combination of two components: (1) measurement error analogous to random error in classical test score theory and (2) the unique component of true score that is different from traits and methods. That is, it is impossible to separate random measurement error (random error variance) from uniqueness (specific variance) due to weak trait and method effects.

The hierarchical confirmatory factor analysis (HCFA) overcomes this limitation (Anderson 1985; Marsh and Hocevar 1988). However, the HCFA model requires rather stringent assumptions and constriants (see Kumar and Dillon 1990 for a detailed discussion of constraints). We feel they are too restrictive to be met in typical consumer research settings, limiting the usefulness of the model. Also, none of the data sets examined herein permits a full illustration of the HCFA model, because it requires at least two measures for each trait-method combination. Given these considerations, the HCFA approach is not included in this paper.

2   Convergence failures are common when method factors are added to trait factors and were found on occasion for these data sets. Nevertheless, by judicious choices of starting values, we were able to achieve complete solutions in all instances. Note that problems of convergence sometimes arise because of overfitting. See discussion of practical relevance and tests thereof in the Discussion where it is suggested that overfitting occurs in two data sets for the CFA analyses.

3   We count as improper solutions only those that are outside the normal range and statistically significant. A number of offending estimates were found in the analyses, consisting of parameter estimates outside the ranges of acceptability. For example, Arora's (1982) data yielded two negative error variances. But because these were statistically nonsignificant they are not counted as improper solutions in Table 2. Instead, they are termed incongruous solutions herein. In either case the results suggested that the model is misspecified or unidentified. We tried one practical way of handling negative error variances, setting error variance estimates to zero (e.g., Dillon, Kumar, and Mulani 1987). This analysis gave quite similar goodness-of-fit results: χ^2 (14) = 16.64, $p < .28$. No offending estimates occurred, but one incongruous estimate (nonsignificant θ_i) was present, raising questions about the adequacy of the model. In sum, the conclusion remained the same; the CFA model was rejected for the Arora data.

4   No standard errors are available for trait correlations in Arora (1982) and Seymour and Lessne (1984) because of the empirical underidentification problem. It should be noted, however, that the DPMs fit poorly in these two data sets.

REFERENCES

Anderson, James C. (1985), "A Measurement Model to Assess Measure-Specific Factors in Multiple-Informant Research," Journal of Marketing Research, 22 (February), 86-92.

Arora, Raj (1982), "Validation of an S-O-R Model for Situation, Enduring, and Response Components of Involvement," Journal of Marketing Research, 19 (November), 505-516.

Bagozzi, Richard P. (1990), "Structural Equation Models in Marketing Research," in The Advanced Research Techniques Forum Proceedings, Chicago: American Marketing Association.

Bentler, Peter and D. Bonett (1980), "Significance Tests and Goodness of Fit in the Analyses of Covariance Structures," Psychological Bulletin, 88 (3), 588-606.

Browne, Michael W. (1984), "The Decomposition of Multitrait-Multimethod Matrices," British Journal of Mathematical and Statistical Psychology, 37 (May), 1-21.

_____ (1985), "MUTMUM: Decomposition of Multitrait-Multimethod Matrices," working paper, Department of Statistics, University of South Africa, Pretoria.

_____ (1989), "Relationships Between an Additive Model and a Multiplicative Model for Multitrait-Multimethod Matrices," in Multiway Data Analysis, eds. R. Coppi and S. Bolasco, Amsterdam: Elsevier, 507-520.

Campbell, Donald T. and Donald W. Fiske (1959), "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," Psychological Bulletin, 56 (March), 81-105.

_____ and Edward J. O'Connell (1967), "Methods Factors in Multitrait-Multimethod Matrices: Multiplicative Rather than Additive?" Multivariate Behavioral Research, 2 (October), 409-426.

_____ and Edward J. O'Connell (1982), "Methods as Diluting Trait Relationships Rather Than Adding Irrelevant Systematic Variance," in Forms of Validity in Research, eds. David Brinberg and Louise A. Kidder, San Francisco: Jossey-Bass, 93-111.

Cote, Joseph A. and M. Ronald Buckley (1987), "Estimating Trait, Method, and Error Variance: Generalizing Across 70 Construct Validation Studies," Journal of Marketing Research, 24 (August), 315-318.

_____ and M. Ronald Buckley (1988), "Measurement Error and Theory Testing in Consumer Research: An Illustration of the Importance of Construct Validation," Journal of Consumer Research, 14 (March), 579-582.

Dillon, William R., Ajith Kumar, and Narendra Mulani (1987), "Offending Estimates in Covariance

    Structure Analysis: Comments on the Causes of and Solutions to Heywood Cases,"

    Psychological Bulletin, 101 (1), 126-135.

Fiske, Donald W. (1982), "Convergent-Discriminant Validation in Measurements and Research

    Strategies," in Forms of validity in Research, David Brinberg and Louis A. Kidder eds., San

    Francisco: Jossey-Bass, 77-92.

Foxman, Ellen R., Patriya S. Tansuhaj, and Karin M. Ekstrom (1989), "Family Members' Perceptions

    of Adolescents' Influence in Family Decision Making," Journal of Consumer Research, 15

    (March), 482-491.

Jackson, Douglas N. (1969), "Multimethod Factor Analysis in the Evaluation of Convergent and

    Discriminant Validity," Psychological Bulletin, 72 (1), 30-49.

Jöreskog, Karl G. and Dag Sörbom (1988), LISREL 7: A Guide to the Program and Applications,

    Chicago: SPSS^X.

Kenny, David A. (1979), Correlation and Causality. New York: Wiley.

Kumar, Ajith and William R. Dillon (1990), "On the Use of Confirmatory Measurement Models in the

    Analysis of Multiple-Informant Reports," Journal of Marketing Research, 28 (February), 102-

    111.

Lastovicka, John L., John P. Murry, Jr., and Eric Joachimsthaler (1990), "Evaluating the

    Measurement Validity of Lifestyle Typologies With Qualitative Measures and Multiplicative

    Factoring," Journal of Marketing Research, 28 (February), 11-23.

Marsh, Herbert W. and Dennis Hocevar (1988), "A New, More Powerful Approach to Multitrait-

    Multimethod Analyses: Application of Second-Order Confirmatory Factor Analysis," Journal of

    Applied Psychology, 73 (1), 107-117.

Maxwell, Albert E. (1977), Multivariate Analysis in Behavioral Research. London: Chapman & Hall.

Menezes, Dennis and Norbert F. Elbert (1979), "Alternative Semantic Scaling Formats for Measuring

    Store Image," Journal of Marketing Research, 16 (February), 80-87.

McDonald, Roderick P. and Herbert W. Marsh (1990), "Choosing a Multivariate Model: Noncentrality

    and Goodness-of-fit," Psychological Bulletin, 107 (March), 247-255.

Mulaik, Stanley A., L. R. James, J. Van Alstine, N. Bennett, and S. Lind, and C. D. Stillwell (1989), "Evaluation of Goodness-of-fit Indices for Structural Equation Models," Psychological Bulletin, 105 (3), 430-445.

Peter, J. Paul (1981), "Construct Validity: A Review of Basic Issues and Marketing Practices," Journal of Marketing Research, 18 (May), 133-145.

Rindskopf, David (1984), "Structural Equation Models: Empirical Identification, Heywood Cases and Related Problems," Sociological Methods & Research, 13 (August), 109-119.

Schmitt, Neal and Daniel M. Stults (1986), "Methodology Review: Analysis of Multitrait-Multimethod Matrices," Applied Psychological Measurement, 10 (March), 1-22.

Searle, Shayle R. (1982), Matrix Algebra Useful for Statistics, New York: Wiley.

Seymour, Daniel and Greg Lessne (1984), "Spousal Conflict Arousal: Scale Development," Journal of Consumer Research, 11 (December), 810-821.

Swain, A. J. (1975), "Analysis of Parametric Structure for Variance Matrices," unpublished dissertation, University of Adelaide.

Werts, Charles E., Karl G. Jöreskog, and Robert L. Linn (1972), "A Multitrait-Multimethod Model for Studying Growth," Educational and Psychological Measurement, 32, 655-678.

Widaman, Keith F. (1985), "Hierarchically Nested Covariance Structure Models for Multitrait-Multimethod Data," Applied Psychological Measurement, 9 (March), 1-26.

Williams, Larry J., Joseph A. Cote, and M. Ronald Buckley (1989), "Lack of Method Variance in Self-Reported Affect and Perceptions at Work: Reality or Artifact?" Journal of Applied Psychology, 74 (3), 462-468.

Wothke, Werner and Michael W. Browne (1990), "The Direct Product Model for the MTMM Matrix Parameterised as a Second Order Factor Analysis Model," Psychometrika, in press.

TABLE 1

SUMMARY OF NESTED CONFIRMATORY FACTOR ANALYSIS TESTS FOR TRAIT AND METHOD EFFECTS

Arora (1982)

Null	Method-Only	
$X^2(36) = 571.70$ $p \cong .00$	$X^2(24) = 271.20$ $p \cong .00$	$X_d^2(12) = 300.50$ $p < .001$
Trait-Only	Trait-Method	
$X^2(24) = 107.62$ $p \cong .00$	$X^2(12) = 15.41$ $p \cong .22$	$X_d^2(12) = 92.21$ $p < .001$
$X_d^2(12) = 464.08$ $p < .001$	$X_d^2(12) = 255.79$ $p < .001$	

Foxman, Tansuhaj, & Ekstrom (1989)

Null	Method-Only	
$X^2(36) = 332.60$ $p < .001$	$X^2(24) = 101.04$ $p < .001$	$X_d^2(12) = 231.56$ $p < .001$
Trait-Only	Trait-Method	
$X^2(24) = 133.96$ $p < .001$	$X^2(12) = 13.28$ $p \cong .35$	$X_d^2(12) = 120.68$ $p < .001$
$X_d^2(12) = 198.64$ $p < .001$	$X_d^2(12) = 87.76$ $p < .001$	

Menzes and Elbert (1979)

Null	Method-Only	
$X^2(36) = 1699.45$ $p < .001$	$X^2(24) = 591.41$ $p < .001$	$X_d^2(12) = 1108.04$ $p < .001$
Trait-Only	Trait-Method	
$X^2(24) = 66.66$ $p < .001$	$X^2(12) = 9.04$ $p \cong .70$	$X_d^2(12) = 57.62$ $p < .001$
$X_d^2(12) = 1632.79$ $p < .001$	$X_d^2(12) = 582.37$ $p < .001$	

Seymour and Lessne (1984)

Null	Method-Only	
$X^2(36) = 650.16$ $p < .001$	$X^2(24) = 270.14$ $p < .001$	$X_d^2(12) = 380.02$ $p < .001$
Trait-Only	Trait-Method	
$X^2(24) = 75.99$ $p < .001$	$X^2(12) = 9.99$ $p \cong .62$	$X_d^2(12) = 66.00$ $p < .001$
$X_d^2(12) = 574.17$ $p < .001$	$X_d^2(12) = 260.15$ $p < .001$	

TABLE 2

SUMMARY OF GOODNESS OF FIT TESTS FOR CONFIRMATORY FACTOR ANALYSES

Study	χ^2(df)	p	AGFI	RMR	Number of Large Standardized Residuals	Other
Arora (1982)	15.41(12)	.22	.87	.06	5	3 nonsignificant θs
Foxman, Tansuhaj, and Ekstrom (1989)	13.28(12)	.35	.94	.03	0	$\Psi_{21} = 1.14(.51)^a$ 1 nonsignificant θ
Menezes and Elbert (1979)	9.04(12)	.70	.97	.02	1	1 nonsignificant θ
Seymour and Lessne (1984)	9.99(12)	.62	.92	.04	1	$\lambda_{55} = -.26(.07)$ $\lambda_{66} = -.41(.18)$ 3 nonsignificant θs

[a]Standard error in parentheses

TABLE 3

SUMMARY OF GOODNESS-OF-FIT TESTS FOR THE DIRECT PRODUCT MODEL

Study	χ^2(df)	p	AGFI	RMR	Goodness-of-fit Diagnostics Other
Arora (1982)	52.12(25)	.00	.82	.15	"$\Gamma_{8,13}$ may not be identified"
Foxman, Tansuhaj, and Ekstrom (1989)	30.04(25)	.22	.93	.06	none
Menezes and Elbert (1979)	26.49(25)	.38	.96	.02	none
Seymour and Lessne (1984)	80.96(25)	.00	.75	.19	"ϕ_{74} may not be identified"

TABLE 4

SUMMARY OF CONVERGENT AND DISCRIMINANT VALIDITY FINDINGS FOR DIRECT PRODUCT MODEL

Arora (1982)

$$\Phi_T = \begin{pmatrix} 1.00 & & \\ .31 & 1.00 & \\ .48 & .38 & 1.00 \end{pmatrix}$$

$$\Phi_M = \begin{pmatrix} 1.00 & & \\ .68 & 1.00 & \\ .79 & .73 & 1.00 \end{pmatrix}$$

Foxman, Tansuhaj, and Ekstrom (1989)

$$\Phi_T = \begin{pmatrix} 1.00 & & \\ .63(.10) & 1.00 & \\ .63(.10) & .79(.10) & 1.00 \end{pmatrix}$$

$$\Phi_M = \begin{pmatrix} 1.00 & & \\ .30 & 1.00 & \\ .18 & .54 & 1.00 \end{pmatrix}$$

Menezes and Elbert (1979)

$$\Phi_T = \begin{pmatrix} 1.00 & & \\ .77(.03) & 1.00 & \\ .45(.05) & .49(.05) & 1.00 \end{pmatrix}$$

$$\Phi_M = \begin{pmatrix} 1.00 & & \\ .91 & 1.00 & \\ .90 & .91 & 1.00 \end{pmatrix}$$

Seymour and Lessne (1984)

$$\Phi_T = \begin{pmatrix} 1.00 & & \\ -.06 & 1.00 & \\ .78 & .30 & 1.00 \end{pmatrix}$$

$$\Phi_M = \begin{pmatrix} 1.00 & & \\ .83 & 1.00 & \\ .99 & .80 & 1.00 \end{pmatrix}$$

TABLE 5

SUMMARY OF FINDINGS ACROSS STUDIES

Method/Criterion	Arora (1982)	Foxman, Tansuhaj, and Ekstrom (1989)	Menezes and Elbert (1979)	Seymour and Lessne (1984)
<u>Confirmatory Factor Analysis</u>				
Model Fit	Reject	Reject	Reject	Reject
Convergent Validity	Pass	Mixed(3/9 Fail)	Mixed(2/9 Fail)	Pass
Discriminant Validity	Pass	Fail	Mixed(1/3 Fail)	Pass
<u>Direct Product Model</u>				
Model Fit	Reject	Accept	Accept	Reject
Convergent Validity	Pass	Mixed(2/3 Fail)	Pass	Pass
Discriminant Validity	Pass*	Fail	Pass	Pass*

NOTE.— The first of three discriminant validity criteria could not be examined for Seymour and Lessne (1984) because standard errors were not available.