Methodology Review: Analysis of Multitrait-Multimethod Matrices

Neal Schmitt
Michigan State University

Daniel M. Stults
Quaker Oats Company

Procedures for analyzing multitrait-multimethod (MTMM) matrices are reviewed. Confirmatory factor analysis (Jöreskog, 1974) is presented as a general model allowing evaluation of the discriminant and convergent validity of MTMM matrices, both as a whole and in individual trait-method units. However, it is noted that this model is deficient with regard to analysis of trait-method interactions of the type described by Campbell and O'Connell (1967, 1982). Composite direct product models described by Browne (1984) are one possible solution to this problem. Further, more systematic use of hypothesis testing regarding convergent and discriminant validity in nested hierarchical models is recommended (Widaman, 1985), as well as the use of a procedure to cross-validate models of MTMM matrices described by Cudeck and Browne (1983).

Classical treatments of measurement error have always been concerned with random error in the measurement of social phenomena. In 1959, Campbell and Fiske drew attention to nonrandom error—that error which serves to increase the intercorrelations of variables because of the proximity in "time, space, or structure" with which they are measured. Campbell and Fiske proposed the use of a multitrait-multimethod matrix (MTMM) to determine the extent of *true* relationship among traits in the presence of both method variance and random error. The purposes of this review are

1. To consider alternatives to the analyses of MTMM matrices in light of this major purpose;
2. To review the literature which has proposed and compared alternative approaches to the analysis of MTMM analyses; and
3. To make some recommendations concerning future analyses of MTMM matrices.

The Campbell-Fiske Criteria

A MTMM matrix consists of the intercorrelations of more than one trait measured by more than one method. The general form of the MTMM matrix is illustrated in Table 1. In this matrix, the correlations on the diagonal (e.g., r_{A1A1}) are internal consistency measures of *reliability*. The triangular sections along the diagonal consisting of correlations among traits measured by a single method (e.g., r_{A1A2}) are referred to as *monomethod triangles*. The values that are underlined are correlations between different measures of the same trait (e.g., r_{A1B1}), referred to collectively as *validity diagonals*. The blocks bordering on the dotted lines containing correlations among traits measured by different methods (e.g., r_{A1B2}) are referred to as *heterotrait-heteromethod blocks* (HTHM).

Campbell and Fiske (1959, pp. 82–83) suggested four criteria by which to evaluate MTMM matrices:

1. The values on the validity diagonal should be statistically significant and large enough to warrant further examination of validity.

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 10, No. 1, March 1986, pp. 1–22

0146-6216/86/010001-22$2.35

2. The values on the validity diagonal should be higher than the HTHM values in the column and row in which the individual validity value is located.
3. The validity value must be higher than the off-diagonal values in its monomethod triangle; that is, a variable should correlate higher with an independent effort to measure the same trait than with other traits measured with the same method.
4. The patterns of trait interrelationship should be the same in all heterotrait triangles in both monomethod and heteromethod blocks.

Early researchers were quick to point out three problems with Campbell and Fiske's proposals for MTMM analysis. First, Campbell and Fiske had provided no method for quantifying the degree to which criteria were met, hence judgments were necessarily of a qualitative nature (Jackson, 1969). Second, and perhaps more significant, various authors pointed to the desirability of separating method variance from random error. Campbell and Fiske indicated that the evaluation of the MTMM correlations must account for the reliabilities of the individual measures. If A_2 in Table 1 is measured with low reliability, then conclusions about the method variance in A_1 and A_3 may be inflated. Jackson (1969) and Althauser and Heberlein (1970) suggested correcting the MTMM matrix for attenuation due to unreliability. The third problem was that the Campbell-Fiske criteria were incomplete. Several authors (Althauser & Heberlein, 1970; Alwin, 1974; Kalleberg & Kluegel, 1975; Krause, 1972) pointed out that implicit in the Campbell-Fiske criteria were the assumptions that there are no correlations between trait and method factors; that all traits are equally influenced by method factors; and that method factors are uncorrelated. As noted below, the problem of intercorrelated trait and method factors has continued to be a vexing one (Campbell & O'Connell, 1967, 1982).

The next section outlines the various methods which have been proposed to deal with some or all of these three problems. However, the review of research in which MTMM matrices were presented indicated that at least half the studies simply reported the matrix of intercorrelations and discussed them in light of the Campbell-Fiske criteria outlined above.

Proposed Analysis Methods

Analysis of Variance of MTMM Matrices

Perhaps the most commonly used method of summarizing a MTMM matrix is by means of an analysis of variance (AOV) paradigm, proposed by

Table 1
MTMM Matrix for Three Traits (1,2,3) and Three Methods (A,B,C)

	Method A			Method B			Method C		
	1	2	3	1	2	3	1	2	3
Method A									
1	r_{A1A1}								
2	r_{A1A2}	r_{A2A2}							
3	r_{A1A3}	r_{A2A3}	r_{A3A3}						
Method B									
1	r_{A1B1}	r_{A2B1}	r_{A3B1}	r_{B1B1}					
2	r_{A1B2}	r_{A2B2}	r_{A3B2}	r_{B1B2}	r_{B2B2}				
3	r_{A1B3}	r_{A2B3}	r_{A3B3}	r_{B1B3}	r_{B2B3}	r_{B3B3}			
Method C									
1	r_{A1C1}	r_{A2C1}	r_{A3C1}	r_{B1C1}	r_{B2C1}	r_{B3C1}	r_{C1C1}		
2	r_{A1C2}	r_{A2C2}	r_{A3C2}	r_{B1C2}	r_{B2C2}	r_{B3C2}	r_{C1C2}	r_{C2C2}	
3	r_{A1C3}	r_{A2C3}	r_{A3C3}	r_{B1C3}	r_{B2C3}	r_{B3C3}	r_{C1C3}	r_{C2C3}	r_{C3C3}

Guilford (1954) and further developed and illustrated by Stanley (1961), Boruch and Wolins (1970), and Boruch, Larkin, Wolins, and MacKinney (1970). In an AOV model, each observed variable is the sum of four components: (1) a general factor that underlies all measures of a person across traits and methods; (2) a trait dimension on which all measures identify the person as being superior or inferior to her/his location on the general factor; (3) a method factor that measures the extent to which a particular measurement method gives higher or lower scores on all traits to a particular person; and (4) random error.

As the AOV model is usually applied (King, Hunter, & Schmidt, 1980), a number of restrictive assumptions are made. Each trait and method factor is defined as independent of the general factor and independent of each other. Note that this independence assumption precludes the estimation of trait intercorrelations, method intercorrelations, or trait-method intercorrelations. Further, because the model is usually applied by averaging correlations of different types (i.e., validity values, heterotrait-monomethod [HTMM] values, monomethod-heterotrait [MMHT] values, and HTHM values), it also assumes that the variances of different trait and method factors are equal. Further, the Stanley (1961) derivations assume that persons, traits, and methods are random factors. While this may be justifiable in the case of persons, it would not be likely for traits and methods in most applied measurement situations.

These rather restrictive assumptions may or may not prove to be a liability in any given application. However, the most serious practical limitation is that the AOV provides only a global estimate of trait, method, and person variance and does not allow the evaluation of individual trait-method units. In some cases (e.g., King et al., 1980), this summary measure is all that the researchers desired, but in most applications the research objectives include an evaluation of individual trait-method units. A further limitation has been noted by Stanley (1961) and King et al. (1980) and was alluded to above: Any potential estimate of trait-method interactions is impossible unless data collection is repeated. In the absence of repeated measures, the interaction of persons, traits, and methods serves as the error term in AOV analyses.

King et al. (1980) presented a good example of the use of the AOV approach. They were interested in evaluating the degree of halo (method bias, in Campbell-Fiske terms) in ratings for a forced-choice rating instrument. Their MTMM matrix consisted of ratings by five raters (methods) on four different performance dimensions (traits). Following the formulations of Stanley (1961), they computed the averages of three different types of correlations in the MTMM matrix:

1. $\bar{r}_g$, which represents the average of correlations in which the trait and rater were different, referred to as HTHM values by Campbell and Fiske (1959);
2. $\bar{r}_M$, which represents the average of the correlations between variables involving the same method, or rater in this instance, referred to as HTMM values by Campbell and Fiske; and
3. $\bar{r}_T$, which represents the average of the correlations between variables involving the same trait, referred to as monotrait-heteromethod (MTHM) values by Campbell and Fiske.

Stanley showed how these values can be related to a three-way analysis of variance, that is, Persons × Traits × Methods. The general factor is analogous to the person effect, the method or halo factor to the Person × Method interaction, and the trait effect to the Person × Trait interaction.

The calculations and results of the applications of this model to one set of data from the King et al. (1980) paper are presented in Table 2. The average HTHM correlation was 29, hence the estimate of the variance due to Persons (due to a general factor) is 29. The average HTMM correlation was 52, so the estimate of variance due to the halo or method factor was 23 (52 − 29). Similarly, the variance attributable to trait differences was computed by subtracting the general component from the average MTHM value (42). The remaining variance is attributable to a combination of the interactions of Persons, Traits, and Methods, and measurement error. When estimates of measurement error are available, these two effects are separable; but in most applications these two sources of variance are confounded. This last variance component is also

Table 2
Example of AOV Analysis of an MTMM Matrix

Effect	Formula Estimate	Estimate of Effect
Persons (General or σ^2_g)	$\bar{r}_g$	29
Method (σ^2_M)	$\bar{r}_M - \bar{r}_g$	23
Trait (σ^2_T)	$\bar{r}_T - \bar{r}_g$	13
Interaction + Error	$1 - \sigma^2_g - \sigma^2_M - \sigma^2_T$	35

Note. Data are Time 1 figures from the King et al. (1980) study. Because they had estimates of the reliability of their findings, King et al. were able to separate the effect of measurement error from the interaction between Persons, Methods, and Traits. In most studies, these two components are not separable and are simply considered error.

used as an error term to test the significance of the Person, Trait, and Method effects (Stanley, 1961).

The AOV approach has been used by researchers in a wide variety of contexts. Most frequent has been its application in studying the "construct validity" of ratings (Lee, Malone, & Greco, 1981; Roberts, Milich, Loney, & Caputo, 1981; Turnage & Muchinsky, 1982). Data collected by measures produced by different types of scaling procedures have also frequently been evaluated by the AOV procedure (Dickinson & Tice, 1977; Dickinson & Zellinger, 1980; Freedman & Stumpf, 1978; Johnson, Smith, & Tucker, 1982; Schriesheim & DeNisi, 1980). Comparisons of different instruments measuring similar traits using the AOV have been made by several investigators (Mayes & Ganster, 1983; Mellon & Crano, 1977; Tinsley & Kass, 1980). Herzberger and Clare (1979) investigated hypotheses concerning attribution theory by comparing observations of actors and observers in various situations using the AOV summary measures.

Nonparametric Analysis of MTMM Matrices: The Generalized Proximity Function

Hubert and Baker (1978, 1979) proposed the nonparametric equivalent of the AOV procedure as a special variant of a generalized proximity function. The procedure begins, as does the AOV, with the computation of three indices: (1) the average of the same-trait correlations, or the validity diagonals in Campbell-Fiske terms; (2) the difference between the average of the same-trait correlations and the average for different traits measured under different methods; and (3) the difference between the average of the same-trait correlations and the average of the same-method correlations. Whereas the AOV approach uses the usual analysis of variance *F*-tests to test the significance of trait and method variance, Hubert and Baker developed nonparametric significance tests. Another difference is that most users of the AOV approach have been primarily interested in the variance accounted for by traits and methods; the Hubert and Baker focus is on the tests of significance.

Hubert and Baker (1978) began with the null hypothesis that the MTMM matrix exhibits no patterning of observations into methods or traits; that is, the actual assignment of the tests to various trait-method combinations was obtained by a random-labeling process. Under this null hypothesis, the expected value of the validity diagonal is equal to the average off-diagonal correlation in the matrix as a whole, and the expected values of the second and third indices cited above are zero. Formulas for the variance of these statistics are also presented. Tests of significance are computed either with an exact test of the probability of a particular patterning or by monte carlo simulation.

The Hubert-Baker approach is, like the AOV approach, an intuitively appealing summary of the entire MTMM matrix. Since their procedure employs nonparametric notions, none of the usual assumptions regarding normality and homogeneity of variance assumptions are necessary. However, problems of differential reliability of measurement and trait-method intercorrelations are not addressed. As with the AOV approach, the indices derived by Hubert and Baker provide information regarding the MTMM matrix as a whole rather than for each trait-method unit. Further, a test of the patterning of trait interrelationships (Campbell & Fiske's fourth criterion) remains to be developed.

Partial Correlations

An hypothesis regarding the failure of some correlational data to meet MTMM criteria is that all data across or within method are affected by leniency (Schriesheim, 1981a, 1981b) or halo error (Holzbach, 1978). Since these factors inflate all correlations in the MTMM matrix, they tend to mask any evidence for discriminant validity which may exist. Both of these authors developed measures of these general "bias" factors and partialled their effects out of the MTMM matrix prior to the assessment of convergent and discriminant validity, using the AOV method described above.

Schriesheim (1981b) examined the convergent and discriminant validity of both the zero-order correlation matrix and a first-order partial MTMM correlation matrix, controlling for leniency as measured by a leniency scale (Schriesheim, 1981a). Controlling for leniency resulted in relatively small decreases in the size of the convergent validities, but also gave fairly significant increases in the degree to which Campbell-Fiske criteria for discriminant validity were met. In performing the comparisons regarding the discriminant validity of the zero-order correlation matrix and the partialled matrix, Schriesheim simply counted the number of times the three conditions for discriminant validity were met. The use of partial correlations, then, aided in identifying the unique trait variance or discriminant validity, while producing little decrement in assessed convergent validity.

Holzbach (1978) used a similar analysis in examining the convergent and discriminant validity of performance ratings supplied by supervisors, peers, and ratees themselves. An overall rating by each of these three groups of raters was partialled out of their ratings on six individual performance dimensions. The assumption was that this global assessment of the individual being rated represented halo.

Holzbach then analyzed both zero-order correlation matrices and partial-correlation matrices using the AOV methods. Removal of the overall effectiveness variance failed to improve discriminant validity overall, but did reduce the halo effect substantially for superior ratings. Intercorrelations among traits for supervisor ratings decreased from a mean of .62 to .12. Applying the same procedure to four previous studies of performance ratings, Holzbach found that controlling for the overall effectiveness ratings resulted in a general lowering of halo and convergent validity indices with little effect on discriminant validity indices; thus discriminant validity improved *relative* to halo and convergent validity.

While Schriesheim's and Holzbach's analyses are empirically correct, they are likely of limited value for several reasons. First, the use of AOV or counting rules on partial correlations contains the same problems as AOV of zero-order MTMM matrices. Second, the use of partial correlations demands that the experimenter know that the focal variables under investigation may be contaminated by some other factor (such as leniency or social desirability), and necessitates the development or availability of an adequate measure of the biasing factor to include in the instrumentation. Third, the use of partial correlation techniques in this fashion implies that all of the general factor variance in a set of ratings is halo error rather than "true" variance. This approach to rating errors has recently generated a small controversy highlighting precisely this problem with the use of partial correlations (Harvey, 1982; Hulin, 1982; Landy, Vance, & Barnes-Farrell, 1982; Murphy, 1982).

Smallest Space Analysis

The Guttman-Lingoes smallest space analysis

(SSA) procedure has been used to assess MTMM matrices in at least one instance (Levin, Montag, & Comrey, 1983). This nonparametric multidimensional scaling (MDS) technique results in the representation of variables in space, the correlations serving as measures of distance. While the procedure produces a goodness-of-fit test regarding the degree to which the spatial representation of the variables is useful in reproducing the correlation matrix, it provides no quantitative detail regarding the degree to which measurements of a given variable are the result of method of measurement or the underlying hypothesized trait. Interpretation of SSA involves a subjective analysis of the degree to which various measures of the same trait are or are not represented in similar parts of a scattergram (Schlesinger & Guttman, 1969).

SSA adds little to the assessment of whether a given MTMM matrix meets the criteria for convergent and discriminant validity. While examination of the representation of variables in space would allow comparative statements about the degree to which a variable seems to be more or less influenced by a configuration of variables representing a method or trait, it does not seem to the present authors that it offers much more than a visual inspection of the correlation matrix would provide.

More generally, as a form of nonmetric MDS, the use of SSA involves the removal of the first general factor from a matrix of correlations (Davison, 1985) in those cases which are likely to occur most frequently. Further, the factors identified from components analysis (if the first factor is not considered) are virtually identical to those arising from MDS solutions. Since a general factor appears prominently in many social science data collection efforts, the use of SSA or other MDS methods will greatly enhance the appearance of discriminant ability. The choice between these techniques then becomes a matter of whether the researcher believes that the general factor is substantively meaningful or represents a response bias or methodological artifact. Treating a general factor as error would be shortsighted if human abilities are to be analyzed, but is more problematic in the area of attitude measurement or performance evaluation. (For a spirited exchange on this issue in the performance evaluation area, see the aforementioned series of papers by Harvey, 1982; Hulin, 1982; Landy, Vance, & Barnes-Farrell, 1982; Murphy, 1982.) MDS methods cannot easily be used to choose between alternative models of a MTMM matrix, and do not allow for estimates of variance attributable to various sources (trait or method or general factors); therefore, it appears that MDS techniques have limited value in the analysis of MTMM matrices.

Exploratory Analyses of MTMM Matrices

There were several early attempts (Golding & Seidman, 1974; Jackson, 1969; Tucker, 1967) to use factor or components analyses to assess convergent and discriminant validity. These methods are exploratory in the sense that a given measure may have loadings on more than one (or all) trait factors, and that one of the objectives of the analyses is to see if some common factor solution can explain measures collected with different methods. These methods were frequently used in early analyses of MTMM matrices (Schmitt et al., 1977), but very little use has been made of exploratory analyses in the last decade. Perhaps the major reason why researchers have abandoned these methods is that they are inconsistent with the Campbell-Fiske criteria and logic. Correlating measures across different methods is done with confirmatory hypotheses in mind. The researcher is not interested in discovering underlying factor structure, but rather in confirming or disconfirming the existence of a single a priori structure across various methods of data collection.

Confirmatory Factor Analytic Model

A confirmatory factor analytic (CFA) model which allows evaluation of each trait-method unit was presented by Werts and Linn (1970), who described it as a special case of analysis of covariance structures (Jöreskog, 1969, 1970). Earlier reviews described the flexibility of the Werts-Linn approach (Alwin, 1974; Schmitt, Coyle, & Saari, 1977), and several examples of its use have been presented (e.g., Kalleberg & Kluegel, 1975; Kenny, 1976; Schmitt, 1978). This section describes the logic of the technique and cites examples of its use in various substantive areas. In a subsequent section CFA is used as a general model by which to evaluate MTMM matrices. Its advantages are highlighted and some remaining problems (Browne, 1984; Widaman, 1985) are described, both in the use of CFA and in the use of MTMM logic in general.

To employ the method suggested by Werts and Linn (1970) and to analyze all potential hypotheses regarding a MTMM matrix, a researcher must have a minimum of three traits measured by three or more methods. While this may appear to be a significant limitation, the model can, with some more restrictive assumptions, be used to test MTMM matrices with smaller dimensions (e.g., Kenny, 1976). Use of other methods applicable to smaller matrices would necessitate the same assumptions whether or not they are explicitly stated (Alwin, 1974).

In CFA analyses of MTMM matrices, the model for each observed variable is comprised of three components: a trait component, a method component, and a random error component. Recall the need for distinguishing between the random error which is the concern of classical measurement theorists and the systematic error or method bias which was the concern of Campbell and Fiske. In the general model, trait and method factors may be correlated while the random error associated with each measured variable is uncorrelated with trait or method factors. A diagram of this model is presented in Figure 1 and the parameters estimated by the LISREL procedure (Jöreskog & Sorbom, 1981) are described in Table 3.

The values in Table 3A are factor loadings; those in Table 3B represent the intercorrelations among trait and method factors (because these are correlations of the underlying or latent factors, they represent correlations between constructs measured without error); and those in Table 3C are the unique or random error components. Values of 0.0 and 1.0 are fixed by the researcher and represent her/his hypotheses regarding the structure of the MTMM matrix. Parameters which are free or estimated based on the observed correlation matrix are numbered consecutively in Table 3, and they total 42. Since there are 45 (9 × 10/2) unique elements in the MTMM matrix, three degrees of freedom are available to test the appropriateness of the specified model.

While this is the smallest MTMM matrix for which a full model can be tested, some restrictive assumptions can be used to allow testing of smaller matrices. For example, an absence of relationship among trait and method factors might be specified, which in the Table 3 model would involve fixing parameters 22–27 and 29–31 at 0.0 (later, it will be shown that these parameters produce estimation problems as well). Alternatively, Alwin's (1974) characterization of the Campbell-Fiske criteria—as indicating a lack of correlation among method factors and among trait and method factors, and an equivalent influence of method factors across traits

Figure 1
Illustration of Hypothesized Determinants of Nine Measured Variables Each Comprised of a Trait, Method, and Random Error Component, All Traits and Methods Interrelated

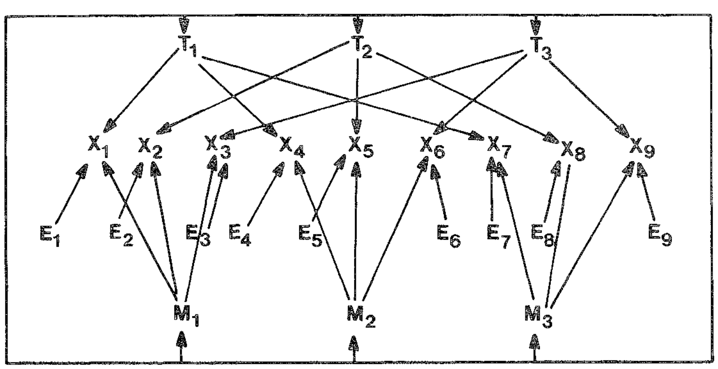


Table 3
Parameters Estimated in Confirmatory Analysis of MTMM
Matrix Depicted in Figure 1

A. Trait and Method Factor Loadings

	T_1	T_2	T_3	M_1	M_2	M_3
X_1	1	0.0	0.0	2	0.0	0.0
X_2	0.0	3	0.0	4	0.0	0.0
X_3	0.0	0.0	5	6	0.0	0.0
X_4	7	0.0	0.0	0.0	8	0.0
X_5	0.0	9	0.0	0.0	10	0.0
X_6	0.0	0.0	11	0.0	12	0.0
X_7	13	0.0	0.0	0.0	0.0	14
X_8	0.0	15	0.0	0.0	0.0	16
X_9	0.0	0.0	17	0.0	0.0	18

B. Intercorrelation of Trait and Method Factors

	T_1	T_2	T_3	M_1	M_2	M_3
T_1	1.0					
T_2	19	1.0				
T_3	20	21	1.0			
M_1	22	23	24	1.0		
M_2	25	26	27	28	1.0	
M_3	29	30	31	32	33	1.0

C. Random Errors Associated with Each Measured Variable

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
34	35	36	37	38	39	40	41	42

Note. X_1 through X_9 indicate the nine measured variables; T_1 though T_3, the three trait factors; M_1 through M_3, the three method factors; 1.0 and 0.0 are fixed values; the numbers from 1 to 42 indicate values estimated by the program based on the observed MTMM matrix.

assessed by that method—would involve fixing parameters 22–33 and specifying a single loading for each method factor. This would gain 18 additional degrees of freedom. Any or all of these restrictive assumptions may be employed to test MTMM matrices with less than three traits or methods if those restrictive assumptions are substantially reasonable.

The CFA approach has been discussed in path analysis terms by Althauser (1974), Althauser and Heberlein (1970), Althauser, Heberlein, and Scott (1971), Alwin (1974), and Werts and Linn (1970); but more frequently the confirmatory nature of the strategy is obvious (Avison, 1978; Bagozzi, 1978, 1980; Jöreskog, 1971, 1974; Kalleberg & Kluegel, 1975; Kenny, 1976; Lee, 1980; Marsh & Hocevar, 1983; Schmitt, 1978; Schmitt, Coyle, & Saari, 1977; Schmitt & Saari, 1978). The content areas in which applications have appeared are varied. Arora (1982) and Bagozzi (1980) outlined the use of the approach in problems related to marketing research. Convergent and discriminant validity of different personality measures was investigated by Watkins and Hattie (1981). The subject of the Schmitt and Saari (1978) application was the perception of leader behavior by self, supervisor, and subordinate sources. A confirmatory analysis of a MTMM matrix consisting of two measures of the self-concept of secondary school students was presented by Marsh and O'Neill (1984) as well as by Marsh and Hocevar (1985), and a confirmatory analysis of college students' responses to two forms of a leisure activities questionnaire was the subject of a paper by Tinsley and Kass (1980). Kalleberg and Kluegel (1975) analyzed responses to different measures of job motivation.

The analysis of attitudinal data presented by Widaman (1985, Table 8) is used as an example of the application of the technique and the type of information it provides. The data were from a study by Kothandapani (1971) in which the affective, behavioral, and cognitive components of attitudes toward churches were assessed using four different attitude scaling approaches. In Table 4 these data are presented in a format that replicates Table 3. As in Table 3, each measured variable loads on a single trait and a single method factor; other loadings were fixed at .00. As can be seen in Table 4 by examining the relative sizes of the trait and factor loadings, several measured variables are heavily influenced by a method factor. This is particularly true of the self ratings. The intercorrelation of the trait and method factors indicates a relatively high intercorrelation between the first and second trait factors, but relatively low correlations with the third trait factor and between all method factors. Trait-method intercorrelations were set at .00. Finally, Table 4C indicates that there are relatively large unique components to some measured variables, particularly those associated with the Guttman technique and self ratings.

The confirmatory approach allows for estimation of parameters, as indicated in Table 4, as well as tests of their significance and the decomposition of each bivariate correlation in the MTMM matrix into a trait and method component. These topics are treated in a later section of this paper in which a general confirmatory model is discussed. While the CFA model has some limitations (Widaman, 1985), the present authors believe it to be the preferred method of analyzing MTMM matrices.

Comparison of Methods of Analyzing MTMM Matrices

Since the Schmitt et al. (1977) review of MTMM matrices, there have been numerous papers that have compared an examination of the MTMM matrix using Campbell-Fiske criteria with one of the methods described above; most frequently used have been the AOV and CFA methods. Four papers have compared alternative methods of analyzing MTMM matrices.

Ray and Heeler (1975) compared (1) restricted maximum likelihood factor analysis (Jöreskog, 1969); (2) a clustering/nonmetric scaling method; and (3) Jackson's (1969) multimethod factor analysis. The MTMM matrix originally presented by Campbell and Fiske (1959) was used as an example. Ray and Heeler pointed out that analyses of this matrix by Jöreskog and by Boruch and Wolins (1970) using the same technique led them to different conclusions. Boruch and Wolins included a general factor while Jöreskog did not. Boruch and Wolins concluded that there were only four distinguishable traits, while Jöreskog retained the five original traits, noting that one ("cheerful") was highly related to other traits. The Jackson (1969) approach to this matrix yielded the conclusion that all traits displayed substantial discriminant validity. The Campbell-Fiske interpretation of this matrix was that there was reason to question the discriminant and convergent validity of "unshakeable poise" (in contrast to Boruch and Wolins who questioned the cheerful trait). The interpretation of the clustering/nonmetric scaling technique was consistent with the Campbell-Fiske interpretation.

Yet another analysis of these same data was conducted by Ray and Heeler (1975). They used a hierarchical clustering method which also indicated four distinctive traits; unshakeable poise was not identified as a separate trait. They found that staff

Table 4
Widaman (1985) Confirmatory Analysis of Multitrait-Multimethod Matrix Presented by Kothandapani (1971)[a]

A. Trait and Method Factor Loadings

	T_1	T_2	T_3	M_1	M_2	M_3	M_4
Thurstone (M_1)							
X_1	.71	.00	.00	.17	.00	.00	.00
X_2	.00	.65	.00	.47	.00	.00	.00
X_3	.00	.00	.71	.75	.00	.00	.00
Likert (M_2)							
X_4	.91	.00	.00	.00	.48	.00	.00
X_5	.00	.69	.00	.00	.73	.00	.00
X_6	.00	.00	.61	.00	.81	.00	.00
Guttman (M_3)							
X_7	.65	.00	.00	.00	.00	.56	.00
X_8	.00	.58	.00	.00	.00	.55	.00
X_9	.00	.00	.74	.00	.00	.58	.00
Self-Rating (M_4)							
X_{10}	.53	.00	.00	.00	.00	.00	.80
X_{11}	.00	.43	.00	.00	.00	.00	.76
X_{12}	.00	.00	.53	.00	.00	.00	.68

B. Intercorrelation of Trait and Method Factors

	T_1	T_2	T_3	M_1	M_2	M_3	M_4
T_1	1.00						
T_2	.61	1.00					
T_3	.04	.24	1.00				
M_1	.00	.00	.00	1.00			
M_2	.00	.00	.00	.17	1.00		
M_3	.00	.00	.00	-.23	.17	1.00	
M_4	.00	.00	.00	.23	.51	.33	1.00

C. Random Errors Associated with Each Measured Variable

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
.71	.57	.00	.18	.16	.13	.50	.67	.52	.38	.54	.49

[a] T_1 through T_3 refer to affective, behavioral, and cognitive components of attitude; X_1 through X_{12} are the measured variables; M_1 through M_4 are the four method factors.

ratings and teammate ratings produced very similar information on the four distinctive trait clusters.

Comparisons of the application of the different techniques and analyses led to the same conclusion with respect to three traits, but the authors disagreed about the distinctiveness of the cheerful and unshakeable poise traits. Ray and Heeler used these results to emphasize the inherent subjectivity of evaluating MTMM matrices even when confirmatory analyses are employed. They recommended the use of multiple techniques and the use of nonmetric data as well as correlations in analyzing the convergence of data collected with different methodologies. They did not, however, suggest how inconsistencies in such multiple analyses should be resolved other than citing the preponderance of evidence from a series of studies. This represents a suggestion that results of MTMM analyses be cross-validated, a topic treated in more depth below.

The Levin et al. (1983) study, discussed above in connection with SSA, also compared smallest space analysis with a varimax-rotated factor analysis of the MTMM matrix. There was substantial agreement between the two procedures as applied to the analysis of four personality constructs measured by the Comrey Personality Scales, the Eysenck Personality Scales and the MMPI. Both procedures led to the conclusion of substantial convergent and discriminant validity. The present authors would not predict similar results; nor would Davison (1985). However, Levin et al. restricted their interpretation to a two-factor solution (even though a three-factor solution was indicated), which was also suggested by the varimax rotation. As in many other comparisons, a series of subjective judgments of similarity or dissimilarity is involved. Levin et al. also indicated that special testing instructions designed to minimize faking and social desirability response sets were provided to respondents. These instructions may have minimized the effect of the general factor.

Lomax (1978) analyzed five MTMM matrices using CFA and Jackson's (1975) revised multimethod factor analysis. He called their use of CFA "exploratory" since four different models of each matrix were evaluated. The five matrices varied in terms of the degree to which they appeared to meet the Campbell-Fiske criteria. Lomax found the greatest agreement between analysis methods when either all or none of the Campbell-Fiske criteria were met. Disagreement across analysis methods was greatest between a matrix presented by Levin (1973), which was judged as having met only the first of the Campbell-Fiske criteria, and a matrix from Roshal, Frieze, and Wood (1971) which supposedly met the first and fourth of the Campbell-Fiske criteria. Lomax and Algina (1979) described the comparisons of exploratory factor analysis and Jackson's multimethod factor analysis for these two matrices. The models evaluated in their "exploratory" analyses included models in which (1) all trait and method factors were correlated zero; (2) correlations among trait factors were estimated, but all others were zero; (3) correlations across trait and method factors were constrained as zero, others were estimated; and (4) all trait-method factors were correlated (which represents the CFA model described in Table 3).

For both sets of data, multimethod factor analysis indicated substantial convergent and discriminant validity. Traits measured by different methods loaded on the same factor, and there was no evidence of a method factor. For the Levin (1973) data, two models seemed to be equally reasonable: the third model described above, in which trait-method correlations are constrained at zero, and another model in which there were three intercorrelated method factors and a single trait factor. Clearly, the latter model is more consistent with the Campbell-Fiske interpretation of the Levin data; namely, that only the first of the Campbell-Fiske criteria were met. For the Roshal et al. (1971) data, neither the third nor the fourth models converged satisfactorily. A fifth model comprised of three method factors and two trait factors (three were intended) seemed to fit reasonably well. Even in this instance, one of the three traits loaded only on method factors. Again, this result is clearly inconsistent with the multimethod factor analysis which seems to highlight, perhaps artifactually, the distinctiveness of traits. In this respect, multimethod factor analysis is similar to the SSA and partial correlation methods described above.

Marsh and Hocevar (1983) compared AOV and CFA analyses of student and self ratings of faculty performance on nine dimensions. They concluded that AOV provides convenient summary and statistical tests of levels of convergent, discriminant, and method effects, but that the AOV effects are not consistent with those given similar labels by Campbell and Fiske. Marsh and Hocevar were not specific as to what these inconsistencies are, but it is important to note that most uses of the AOV approach are not themselves consistent with the approach as outlined by Stanley (1961). Compare, for example, King et al. (1980) and Stanley with Boruch et al. (1970) and Dickinson and Zellinger (1980). To the present authors' knowledge, only the King et al. (1980) use of the AOV approach was consistent with Stanley's original formulation which, on a *matrix* level, seems consistent with the Campbell-Fiske criteria. Subsequent users of the AOV approach would be well-advised to return to the Stanley paper or to King et al. Marsh and Hocevar (1983) do correctly point out that the AOV approach does not allow analysis of trait-method interactions due to the use of the three-way interaction as an error term in the AOV. As stated above, both Stanley (1961) and King et al. (1980) recommended that the measures be replicated for each case within a given study, thus providing independent estimates of the three-way interaction (the method-trait correlations) and the random error term.

The Marsh-Hocevar (1983) preference was clearly CFA. They cited as significant advantages (1) the ability to test various alternative models of the MTMM matrix; (2) the use of underlying variables as opposed to observed variables, which circumvents the reliability problem; and (3) the ability to separate trait, method, and unique variance in each measure. They also mentioned the utility of testing alternate models nested within a general model, though their own analyses did not follow this format.

Comparisons of methods for the analysis of MTMM matrices lead to several conclusions. First, the methods yield different conclusions when the matrix as a whole or the individual traits being measured are marginal in terms of the degree to which they meet Campbell-Fiske criteria. When the Campbell-Fiske criteria are relatively unambiguously met or when there is no evidence of convergent or discriminant validity, then most methods of analyzing MTMM matrices yield identical conclusions. When there is conflicting or inconsistent evidence regarding the degree to which Campbell-Fiske criteria are met, then researchers using different methods of analysis are most likely to reach different conclusions. Second, the major difference among methods involves the degree to which conclusions regarding individual traits can be drawn, in addition to conclusions regarding the matrix as a whole. As an example, CFA allows both types of conclusions, whereas an AOV approach provides analysis only of the matrix as a whole. Third, even those methods which are confirmatory in nature allow for a great deal of experimenter subjectivity in deciding which model of the matrix is appropriate and which criteria for construct validity are met. This is not necessarily bad; one experimenter working in a given situation may realize that discriminant or convergent validity is more important than another experimenter whose research interests are different. However, differences in conclusions may also result from inefficient use of the power of confirmatory analysis and a relatively nonsystematic use of these procedures in analyzing MTMM matrices. For example, the use of nested models to test specific hypotheses about convergent and divergent validity has been described (Jöreskog, 1969, 1974), but frequently illustrations of the technique have presented statistics regarding a variety of models with little or no guidance as to the implications to be drawn from these presentations (Marsh & Hocevar, 1983; Schmitt, 1978).

Summary

Throughout the discussion of various methods of analyzing MTMM matrices, two major questions concerning the convergent and discriminant validity of a set of measures have recurred. Specifically,

1. In the MTMM matrix as a whole, does the technique allow for the evaluation of the importance of
   a. trait-method interactions,

   b. method intercorrelations,
   c. trait versus method factors?
2. Does the technique allow for the separation of trait, method, and random variance in individual trait-method units (i.e., measured variables)?

These major questions subsume the Campbell-Fiske (1959) criteria. As is true of the use of all statistical techniques, researchers have been interested in the *estimation* of the relative size of parameters relating to trait and method factors. They have also been interested in testing hypotheses about traits, trait interrelationships, and trait-method relationships. In the last section of this paper, the use of the CFA approach is presented as the most comprehensive means of estimating and testing hypotheses related to the two major questions cited above.

Confirmatory Factor Analysis as a General Model to Evaluate MTMM Matrices

The CFA model posits that the MTMM matrix Σ can be expressed as a function of common factors as follows:

$$\Sigma = \Lambda\Phi\Lambda' + \Psi \quad , \qquad (1)$$

where Λ is the matrix of factor loadings (see Table 3A),
Φ is the matrix of correlations among factors (see Table 3B), and
Ψ is a diagonal matrix of unique factor variances (see Table 3C).

As can be seen in Table 3, each measured variable is expressed as a function of a trait factor, a method factor, and some unique variance. Jöreskog (1969, 1971), assuming that the factor analytic model expressed in Equation 1 is appropriate and assuming a multivariate normal distribution, developed maximum likelihood estimation procedures for all parameters in Λ, Φ, and Ψ. The following sections outline how CFA allows for hypothesis testing as well as estimating the size of method, trait, and unique variance in providing answers to the two questions posed at the end of the previous section.

Hypothesis Testing

In specifying a model of a MTMM matrix (as in Table 3), a null hypothesis is expressed in each case in which a parameter is left free to be estimated. The significance of each of these parameters is tested by calculating a *z*-ratio: the parameter estimate divided by its asymptotic standard error. Those estimates with absolute values greater than 2.00 are considered significant beyond the .05 level. In addition to these tests of the significance of individual parameter estimates, there is also a χ^2 test of the overall fit of the hypothesized model. This χ^2 test involves a test of the significance of the difference between the observed correlation matrix (if the model were perfect it would completely account for these observed correlations) and the reproduced correlation matrix. The reproduced matrix is the one implied by the parameter estimates; the mechanics of reproducing the matrix are detailed below. The number of degrees of freedom associated with this χ^2 test is equal to the number of independent elements in the correlation matrix minus the number of parameters in the hypothesized model. If the χ^2 is significant, there is basis for rejecting the model and evaluating a model that includes more parameters.

Since this χ^2 test is dependent on sample size, models with relatively good fit may be rejected when sample size is large. Consequently, there have been recommendations that the theoretical and conceptual appropriateness of the model be considered (Browne, 1984, p. 153), as well as the χ^2/df ratio (Jöreskog & Sorbom, 1981). The latter is then taken as a goodness-of-fit measure. In addition, Bentler and Bonett (1980) provided two goodness-of-fit measures. These two practical tests of significance were labeled rho and delta. Rho is a relative index of the degree of off-diagonal covariation among the observed variables which is explained by the model. Specifically, rho is expressed as follows:

$$\rho = [(\chi^2_N/df_N) - (\chi^2_S/df_S)]/[(\chi^2_N/df_N) - 1] \qquad (2)$$

where N refers to a model with no hypothesized relationships, or a null model, and S refers to the model being evaluated. The goodness-of-fit of a model, then, is relative to the degrees of freedom. Delta is an absolute measure of fit in that it represents the proportion of off-diagonal covariation accounted for in a model independent of the degrees of freedom. In formula terms,

$$\Delta = (\chi^2_N - \chi^2_S)/\chi^2_N \quad . \qquad (3)$$

Similar indices have been incorporated in recent versions of LISREL (Jöreskog & Sorbom, 1981), the computer program usually used to obtain CFA. A conventional rule of thumb is that ρ and Δ values should exceed .90 to permit the conclusion that a model is acceptable (Bentler & Bonett, 1980; Tucker & Lewis, 1973; Widaman, 1985).

These various goodness-of-fit indices and tests of significance then provide tests of both the overall model of a MTMM matrix plus tests of individual variables, trait, method, and unique variance components. Another set of questions normally addressed in evaluating a MTMM model is whether all hypothesized method factors, trait factors, or method and trait factor relationships need be included in a model. For example, a researcher, after looking at the results presented in Table 4, might decide to compare that model with a model that includes only two trait factors because T_1 and T_2 were correlated .61. The CFA approach provides a means of testing the significance of the difference of these two models, provided one model is nested within the other. Recently, Widaman (1985) provided a critique of earlier applications of these tests and provided a systematic array of structural models.

Widaman (1985) proposed to first identify the model of best fit and then proceed to make comparisons of this best-fitting model with other models nested within the best-fitting model, as a means of testing various aspects of convergent and discriminant validity. In systematically generating an array of structural models, Widaman considered the relationships among underlying trait factors and method factors separately (in Table 3, the trait factor interrelationships are measured by parameters 28, 32, and 33). In each case, there are three possible structures: (1) no trait (or method) factors; (2) trait (or method) factors with fixed intercorrelations: either 0, indicating a high level of discriminant validity, or 1, indicating a total lack of discriminant validity; and (3) trait (or method) factors with freely estimated intercorrelations. The major contribution of the Widaman critique of the use of CFA was the observation that past applications of the technique did not use it to full advantage in testing hypotheses and evaluating the practical implications or degree of fit of substantively important models. His proposal for a systematic approach to confirmatory analyses of MTMM matrices is briefly outlined above. Widaman's 3C model is apparently the least restrictive model that is identifiable, and would include estimation of all parameters listed in Table 3 above with the exception of parameters 22–24, 25–27, and 29–31, which represent the trait-method intercorrelations. The latter are fixed at zero. The following restrictions provide tests of convergent validity, discriminant validity, and method bias for the matrix as a whole.

1. Comparison of this model with one in which only correlated methods factors were present would provide a test of convergent validity. The difference in the χ^2 tests of fit associated with these two models is itself distributed as a χ^2 with degrees of freedom equal to the difference in degrees of freedom associated with the two models. In terms of the model described in Table 3, the three trait factors would be dropped. There would be no estimation of trait factor loadings (of which there are 9), or of trait intercorrelations (of which there are 3). Hence, in this case, the test of convergent validity has twelve degrees of freedom.
2. Comparison of the full model with a model that includes perfectly intercorrelated trait factors yields a test of discriminant validity for the matrix as a whole. This comparison would have three degrees of freedom for the matrix presented in Table 3, as parameters 19–21 would be fixed at 1.0.
3. An overall test of method bias would be provided by comparing the full model to one which includes no method factors. As in the test for convergent validity, this comparison would yield 12 degrees of freedom as the method factors, their loadings, and their intercorrelations are eliminated.

Comparison of these models is facilitated by the χ^2 tests, but perhaps most importantly by the indices of degree of fit provided by Bentler and Bonett (1980). While these three tests would be of major initial interest to most researchers, additional tests will almost always be suggested by the initial analyses. For example, suppose the first comparison above results in a conclusion that significant convergent validity exists (that is, a model with no trait factors does not fit the data well compared to the full model). However, examination of the trait intercorrelations in the full model indicates that the first two underlying traits are highly correlated. A reasonable modification, then, might be a model with two trait factors. There are obviously many different possibilities, but these possibilities should be explored in a systematic manner; Widaman (1985) offers such a paradigm for organizing the investigation of MTMM matrices.

Widaman's suggestion that the hypothesis testing stage begin after identification of a reasonable model, as opposed to comparisons with the least restrictive model, may also be helpful. Browne (1984) noted that models providing good fit to a particular data set often include what he terms "wastebasket" parameters. He cited two instances: the Jöreskog (1974) analysis of the Campbell-Fiske data, and the Schmitt (1978) analysis of the Ostrom (1969) data. To avoid the inclusion of these parameters, Browne encouraged an examination of the practical meaningfulness of parameters in models even if they fit well, and the use of an empirical cross-validation of models of MTMM matrices.

Obviously, as various hypotheses regarding a MTMM matrix are evaluated, the procedure becomes more exploratory than confirmatory; in fact, Ray and Heeler (1975) titled their use of the confirmatory procedures described here "exploratory." A best-fitting model discovered by this process should be confirmed with additional independent data collection efforts or the cross-validation procedure suggested by Cudeck and Browne (1983). This cross-validation procedure represents a new development and is discussed more fully below.

Estimation of Trait and Method Effects

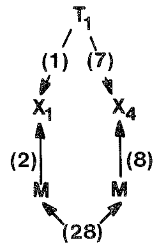
As indicated above, parameters are estimated via maximum likelihood procedures detailed by Jöreskog (1969, 1971). The results of CFA provide the data by which to evaluate the size of the effect of hypothesized trait, method, and unique factors. Using the parameter estimates, such as those presented in Table 4, each bivariate correlation can be decomposed.

Widaman (1985) alluded to the advantages of CFA in the decomposition of the variance of individual measured variables and to the decomposition of individual correlations in a MTMM matrix. These advantages provide significant information to a researcher interested in assessing particular measured variables. It is important to repeat that the decompositions suggested here assume an absence of trait-method correlations. Suppose a researcher is interested in assessing the degree to which individual differences in the first measured variable (X_1) in Table 3 are due to trait, method, and unique variance. The trait variance would be estimated by squaring the first parameter in Table 3 (the factor loading of X_1 on T_1). The method variance would be estimated by squaring parameter 2 (the factor loading of X_1 on M_1). The unique variance associated with X_1 would be estimated by squaring parameter 34 (without trait-method intercorrelations, this would be parameter 25). Similar decompositions of the variance of each measured variable would allow an investigator to draw conclusions about individual measures in the MTMM matrix.

Individual correlations between measured variables can also be decomposed to provide assessment of the degree to which correlations result from common trait or method variance. Schmitt (1978) provided examples of the decomposition of MTHM correlations, HTMM correlations, and HTHM correlations, but his examples included provision for trait-method correlations. Figure 2 provides examples of such decomposition for the three types of correlations in a MTMM matrix. Correlations between measured variables are the sums of the products of the parameters associated with hypothetical paths connecting the two measured variables. So, for a HTHM correlation, the intercorrelation due to common trait variance is a product of the factor loadings of the measured variables on the traits involved and the intercorrelation of the traits. Similarly, the intercorrelation due to common method variance is a product of the factor loadings of each measured variable on appropriate method factors and the intercorrelation of the two method factors. The sum of the common trait and method correlation is the value of the HTHM correlation suggested by the model representing the MTMM matrix.

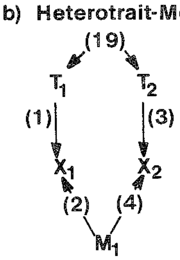
Figure 2
Examples of the Decomposition of Monotrait-Heteromethod Correlations,
Heterotrait-Monomethod Correlations, and Heterotrait-Heteromethod Correlations
(Numbers in Parentheses in the Diagrams are the Parameter Estimates as Designated in Table 2)

(a) Monotrait-Heteromethod Decomposition



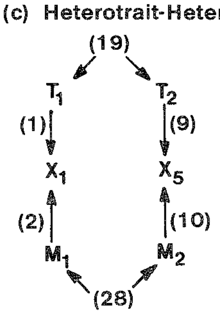
$r_{X_1X_4}$ = (1 x 7) Common Trait Variance
+(2 x 8 x 28) Common Method Variance

(b) Heterotrait-Monomethod Decomposition



$r_{X_1X_2}$ = (1 x 19 x 3) Common Trait Variance
+(2 x 4) Common Method Variance

(c) Heterotrait-Heteromethod Decomposition



$r_{X_1X_5}$ = (1 x 19 x 9) Common Trait Variance
+(2 x 28 x 10) Common Method Variance

Applying these formulas to the data presented in Table 4 and reproducing the correlation between X_1 and X_4 would yield the following:

$$r_{X_1X_4} = (.71 \times .91) + (.17 \times .48 \times .17)$$
$$= .646 + .014$$
$$= .66$$

The fact that X_1 and X_4 index the same trait accounted for most (.646) of the reproduced correlation between these two variables, while method variance accounted for a small portion (.014). The observed correlation in this instance (see Kothandapani, 1971) was .58, so the model overestimated the correlation by .08. Variance in X_1 itself is primarily due to the trait factor ($.71^2 = 50\%$) and the uniqueness associated with this measured variable ($.70^2 = 49\%$), while method variance played a minimal role ($.17^2 = 3\%$). Given that the latter three are independent sources of variance, they should equal 100%; in this case, rounding error produced a slightly larger figure. Given the appropriateness of the model, estimates of the influence of trait, method, and unique factors are available for each measured variable and for each correlation in the MTMM matrix.

Remaining Problems and New Developments

In terms of the general model outlined at the beginning of this paper, CFA as operationalized via LISREL (Jöreskog & Sorbom, 1981) has one deficiency: Trait-method correlations cannot be evaluated. One potential solution to this problem has been developed by Browne (1984). A second problem with CFA is that the search for an appropriate model of a MTMM matrix means that CFA, as employed in most instances, becomes more exploratory than confirmatory. A cross-validation of MTMM models has recently been suggested by Cudeck and Browne (1983). A final problem is that tests of the significance of some parts of a model may be performed in two different ways (against a null hypothesis of $r = .00$ or $r = 1.00$), both of which may be of minimal interest to researchers interested in the actual population value. In this respect, meta-analyses of the substantial MTMM literature may be helpful.

Trait-method correlations. One important difference between Widaman's (1985) proposal and the outline of the confirmatory model given in Table 3 and Figure 1 is that Widaman does not propose the estimation of correlations among trait and method factors. Widaman cites personal communication with Jöreskog and Bentler in making the case that these intercorrelations present logical and empirical problems. Support for the contention that estimation of these parameters produces an identification problem is evident in various applications of CFA using LISREL (Kalleberg & Kluegel, 1975; Lee, 1980; Marsh & Hocevar, 1983; Watkins & Hattie, 1981). These applications have resulted in factor loadings which exceed 1.00; the presence of Heywood cases, in which estimates of unique variance are near zero (indicating total lack of measurement error); and very large standard errors for some parameters resulting from the high intercorrelation of the parameter estimates.

The possibility that trait-method intercorrelations exist was first suggested by Campbell and O'Connell (1967). They presented evidence that method factors may interact in a specific multiplicative way with trait factors: The higher the basic relationship between two traits, the more that relationship is increased when the method is shared. On the other hand, when two traits are basically independent, their correlation, even when measured by the same method, is still zero. Returning to this problem, Campbell and O'Connell (1982) used autoregressive time series models to investigate the hypothesis that differences in method *attenuate* relationships which are more clearly evident when method is held constant than when cross-method correlations are examined. Traditional belief is that some method correlations are inflated above true values more appropriately shown in cross-method comparisons. Campbell and O'Connell provide evidence for a trait-method interaction, and indicate that support for an augmentation model is best. That is, correlated error or method bias tends to exaggerate correlations between the more highly correlated traits. The major point, however, is that there is evidence that method-trait correlations exist.

Recently, Browne (1984) made a similar point concerning existing models of MTMM matrices and proposed a new "composite direct product model" with a multiplicative property. The usual factor analysis model is an additive one; Browne posited that the appropriate expression of a covariance matrix of observed measures is the *product* of the method factors and the trait factors. Browne's hypothesis is that methods act to enhance or dampen the expression of particular traits. The formulas for estimation of parameters in this composite direct product model are in Browne (1984, pp. 13–14). He presented an analysis of the Campbell-Fiske (1959) Table 12, which resulted in a reasonable and interpretable summary of the data in that table, even though a significance test indicated a lack of fit. By contrast, Browne pointed to the results of Jöreskog's (1974) analysis of the same data, in which inclusion of method factors was necessary to provide a reasonably fitting model even though parameters associated with these method factors made little interpretive sense. Browne also pointed out that his direct product model is mathematically equivalent to Tucker's (1966) three-mode factor analysis. Interestingly, since the Schmitt et al. (1977) presentation of methods of analyzing MTMM matrices, the authors are aware of no further use of

Tucker's multimode analysis for this purpose. Future widespread use of Browne's composite direct product model is, of course, dependent on the availability of computer software.

Widaman (1985) has clearly pointed to a potentially important deficiency in the analysis of MTMM matrices. More empirical work regarding the practical importance of trait-method interactions and trait-method correlations seems warranted. Finally, it should be noted, as it has been elsewhere (e.g., Marsh & Hocevar, 1983; Schmitt, 1978; Widaman, 1985), that assuming trait-method correlations (if this represents a reasonable position) to be zero facilitates the interpretation of MTMM matrices. This, of course, was true of the calculations presented above as well. Each measured variable can be represented as a function of three independent sources of variance: trait, method, and random variance.

Cross-validation of MTMM models. As indicated above, most modeling attempts involve considerable trimming and respecifying before an adequate fit is produced. This makes the whole process exploratory and necessitates that attention be directed to cross-validation of these models. Cudeck and Browne (1983) have developed a technique to accomplish cross-validation of competing models. This cross-validation of models of the MTMM matrix involves the following steps: (1) computation of covariance matrices from two samples, a calibration and cross-validation sample; (2) estimation of model parameters in a calibration sample; (3) computation of the reproduced covariance matrix based on these parameter estimates; and (4) comparison of the reproduced covariance matrix with the cross-validation sample covariance matrix.

The reproduced covariance matrix is obtained by minimizing the discrepancy function associated with the maximum Wishart likelihood estimates (see Browne, 1984, Equation 6.3; or Schmitt, 1978, Equation 2). The cross-validation index is computed using the sample covariance matrix from the cross-validation sample and the reproduced covariance matrix from the calibration sample, and computing the discrepancy function. This process is equivalent to cross-validation of linear regression, in which regression weights are obtained in a calibration sample by minimizing the sum of squares as a discrepancy function.

The cross-validation is carried out by evaluating the residual sum of squares in the cross-validation sample using the calibration sample weights. When there is no "shrinkage" from calibration sample to cross-validation sample, an identical residual sum of squares is expected. The degree to which the residual sum of squares is larger in the cross-validation sample than it was in the calibration sample represents the degree to which a regression equation represents a poor model of the linear relationship between predictors and criterion.

Tests of significance versus estimates of relationships. Some tests of significance for individual parameters which are possible using CFA may not be particularly important. For example, if a researcher is interested in the relationships among the trait factors (see Table 3B), he/she is asking questions concerning discriminant validity. Widaman (1985), among others, points out that two "tests of significance" are possible in this instance. First, the hypothesis that the estimates of intercorrelation between underlying factors are significantly greater than zero can be tested. This is consistent with the assertion that a conclusion of discriminant validity means that all trait relationships are nonsignificantly different from zero. Alternatively, the hypothesis that all trait interrelationships are nonsignificantly different from 1.00 can be evaluated. It may be concluded that discriminant validity exists when estimates of the trait correlations were two or more standard errors below 1.0.

Most researchers, however, are more interested in the degree of relationship between underlying traits than in tests of significance. For at least some trait interrelationships, as well as relationships between different methods of data collection, it may be appropriate to resort to meta-analyses (Hunter, Schmidt, & Jackson, 1982) of existing data to generate estimates of these parameters. Relationships between affective, behavior, and cognitive components of attitude have likely been investigated in dozens of studies. Cumulating these studies may result in reasonably good estimates of these interrelationships. These estimates could then be used as fixed values in subsequent research, or they could be used as the value against which to test significance of a particular sample relationship.

King et al. (1980) provided an example of such a meta-analysis. They were interested in the degree of trait, method or halo, and general factor variance in performance ratings, and reviewed the results of eleven different studies which provided multitrait-multirater matrices. They reanalyzed the data from these studies using the AOV approach and found that, on the average, 30% of the variance in ratings was accounted for by halo (which was their primary interest).

Summary and Conclusions

Various procedures (primarily those presented in the last decade) designed to analyze MTMM matrices have been reviewed. Since a similar review in 1977 (Schmitt et al., 1977), three new approaches to the analysis of MTMM matrices have been presented: smallest space analysis (Levin et al., 1983); a nonparametric approach (Hubert & Baker, 1978, 1979); and a composite direct product model (Browne, 1984). Most frequently used in the analyses of MTMM matrices have been the AOV approach (Stanley, 1961) and the CFA approach suggested by Jöreskog (1974). Earlier exploratory factor analysis techniques (Golding & Seidman, 1974; Jackson, 1969, 1975; Tucker, 1967) seem to have been discarded.

The present authors have a strong preference for the analysis of MTMM matrices by the use of CFA. Explaining MTMM matrices on the basis of underlying factors solves the differential reliability of measurement problem cited by early critics of Campbell-Fiske criteria. This approach also allows for the evaluation of convergent and discriminant validity hypotheses at both the matrix and individual measure level. Further, it provides both statistical tests of hypotheses (when hierarchically nested models are used) and goodness-of-fit indices. Use of CFA should be greatly facilitated by Widaman's (1985) paper in which an array of hierarchically nested models is described.

A technique used to compute a cross-validation index for models of MTMM matrices was described and should prove useful. In addition, meta-analyses of some frequently investigated relationships should be conducted so as to provide more definitive answers concerning the convergent and discriminant validity of widely investigated relationships.

At least one significant problem in the analysis of MTMM matrices is not addressed by the CFA approach: evaluation of trait-method interactions (Campbell & O'Connell, 1967). A recent proposal by Browne (1984) to use models he termed composite direct product models may be useful in this regard, as may be the Tucker (1966) multimode factor analysis procedure. Probably because the analytical tools to address the problem have not been available, not much is known about the empirical nature or practical importance of these interactions. Recent work on this problem by Campbell and O'Connell (1982) resulted in the tentative conclusion that correlated error exaggerates the correlation between highly correlated traits relative to traits whose correlation is comparatively low.

Future research regarding the existence and importance of the trait-method interaction and trait-method intercorrelation problems may benefit from the use of monte carlo simulations. Various methods of analysis of MTMM matrices with known structure could be compared on the basis of whether or not appropriate conclusions were drawn.

References

Althauser, R. P. (1974). Inferring validity from the multitrait-multimethod matrix: Another assessment. In H. L. Costner (Ed.), *Sociological Methodology 1973–1974* (pp. 106–107). San Francisco: Jossey-Bass.

Althauser, R. P., & Heberlein, T. A. (1970). Validity and the multitrait-multimethod matrix. In E. F. Borgatta & F. W. Bohrnstedt (Eds.), *Sociological Methodology* (pp. 151–169). San Francisco: Jossey-Bass.

Althauser, R. P., Heberlein, T. A., & Scott, R. A. (1971). A causal assessment of validity: The augmented multitrait-multimethod matrix. In H. M. Blalock, Jr. (Ed.), *Causal models in the social sciences* (pp. 374–399). Chicago: Aldine.

Alwin, D. (1974). An analytic comparison of four approaches to the interpretation of relationships in the multitrait-multimethod matrix. In H. L. Costner (Ed.), *Sociological Methodology 1973–1974* (pp. 79–105). San Francisco: Jossey-Bass.

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

Avison, W. R. (1978). Auxiliary theory and multitrait-multimethod validation: A review of two approaches. *Applied Psychological Measurement, 2,* 433–449.

Bagozzi, R. P. (1978). The construct validity of the affective, behavioral, and cognitive components of attitude by analysis of covariance structures. *Multivariate Behavioral Research, 13,* 9–31.

Bagozzi, R. P. (1980). *Causal models in marketing.* New York: Wiley.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

Boruch, R. F., Larkin, J. D., Wolins, L., & MacKinney, A. C. (1970). Alternative methods of analysis: Multitrait-multimethod data. *Educational and Psychological Measurement, 30,* 833–853.

Boruch, R. F., & Wolins, L. (1970). A procedure for estimation of trait, method, and error variance attributable to a measure. *Educational and Psychological Measurement, 30,* 547–574.

Browne, M. W. (1984). The decomposition of multitrait-multimethod matrices. *British Journal of Mathematical and Statistical Psychology, 37,* 1–21.

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

Campbell, D. T., & O'Connell, E. J. (1967). Method factors in multitrait-multimethod matrices: Multiplicative rather than additive? *Multivariate Behavioral Research, 2,* 409–426.

Campbell, D. T., & O'Connell, E. J. (1982). Methods as diluting trait relationships rather than adding irrelevant systematic variance. In D. Brinberg & L. Kidder (Eds.), *New directions for methodology of social and behavioral science: Forms of validity in research* (pp. 93–111). San Francisco: Jossey-Bass.

Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research, 18,* 147–167.

Davison, M. L. (1985). Multidimensional scaling versus components analysis of test intercorrelations. *Psychological Bulletin, 97,* 94–105.

Dickinson, T. L., & Tice, T. E. (1977). The discriminant validity of scales developed by retranslation. *Personnel Psychology, 30,* 217–228.

Dickinson, T. L., & Zellinger, P. M. (1980). A comparison of the behaviorally anchored rating and mixed standard scale formats. *Journal of Applied Psychology, 65,* 147–154.

Freedman, R. D., & Stumpf, S. A. (1978). Student evaluations of courses and faculty based on a perceived learning criterion: Scale construction, validation, and comparison of results. *Applied Psychological Measurement, 2,* 189–202.

Golding, S. L., & Seidman, E. (1974). Analysis of multitrait-multimethod matrices: A two step principal components procedure. *Multivariate Behavioral Research, 9,* 479–496.

Guilford, J. P. (1954). *Psychometric methods.* New York: Wiley.

Harvey, R. J. (1982). The future of partial correlation as a means to reduce halo in performance ratings. *Journal of Applied Psychology, 67,* 171–176.

Herzberger, S. D., & Clare, G. L. (1979). Actor and observer attributions in a multitrait-multimethod matrix. *Journal of Research in Personality, 13,* 1–15.

Holzbach, R. L. (1978). Rater bias in performance ratings: Superior, self- and peer ratings. *Journal of Applied Psychology, 63,* 579–588.

Hubert, L. J., & Baker, F. B. (1978). Analyzing the multitrait-multimethod matrix. *Multivariate Behavioral Research, 13,* 163–179.

Hubert, L. J., & Baker, F. B. (1979). A note on analyzing the multitrait-multimethod matrix: An application of a generalized proximity function comparison. *British Journal of Mathematical and Statistical Psychology, 32,* 179–184.

Hulin, C. L. (1982). Some reflections on general performance dimensions and halo rating error. *Journal of Applied Psychology, 67,* 165–170.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies.* Beverly Hills CA: Sage.

Jackson, D. N. (1969). Multimethod factor analysis in the evaluation of convergent and discriminant validity. *Psychological Bulletin, 72,* 30–49.

Jackson, D. N. (1975). Multimethod factor analysis: A reformulation. *Multivariate Behavioral Research, 10,* 439–460.

Johnson, S. M., Smith, P. C., & Tucker, S. M. (1982). Response format of the Job Descriptive Index: Assessment of reliability and validity by the multitrait-multimethod matrix. *Journal of Applied Psychology, 67,* 500–505.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183–202.

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika, 57,* 239–251.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36,* 106–133.

Jöreskog, K. G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In R. C. Atkinson, D. H. Krantz, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2, pp. 1–56). San Francisco: W. H. Freeman.

Jöreskog, K. G., & Sorbom, D. (1981). *LISREL V: Analysis of linear structural relationships by maximum likelihood and least squares methods*. Uppsala, Sweden: University of Uppsala.

Kalleberg, A. L., & Kluegel, J. R. (1975). Analysis of the multitrait-multimethod matrix: Some limitations and an alternative. *Journal of Applied Psychology, 60,* 1–9.

Kenny, D. A. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology, 12,* 247–252.

King, L., 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.

Krause, M. S. (1972). The implications of convergent and discriminant validity data for instrument validation. *Psychometrika, 37,* 179–186.

Landy, F. J., Vance, R. J., & Barnes-Farrell, J. (1982). Statistical control of halo: A response. *Journal of Applied Psychology, 67,* 177–180.

Lee, R., Malone, M., & Greco, S. (1981). Multitrait-multirater analysis of performance ratings for law enforcement personnel. *Journal of Applied Psychology, 66,* 625–632.

Lee, S. Y. (1980). Estimation of covariance structure models with parameters subject to functional constraints. *Psychometrika, 45,* 309–324.

Levin, J. (1973). Bifactor analysis of a multitrait-multimethod matrix of leadership criteria in small groups. *Journal of Social Psychology, 89,* 295–299.

Levin, J., Montag, I., & Comrey, A. L. (1983). Comparison of multitrait-multimethod, factor, and smallest space analysis on personality scale data. *Psychological Reports, 53,* 591–596.

Lomax, R. G. (1978). *The multitrait-multimethod matrix: A comparison of alternative analytical procedures*. Unpublished master's thesis, University of Pittsburgh, Pittsburgh, PA.

Lomax, R. G., & Algina, J. (1979). Comparison of two procedures for analyzing multitrait-multimethod matrices. *Journal of Educational Measurement, 20,* 231–248.

Marsh, H. W., & Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. *Journal of Educational Measurement, 20,* 231–248.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First and higher order factor models and their invariance across groups. *Psychological Bulletin, 97,* 562–582.

Marsh, H. W., & O'Neill, R. (1984). Self description questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement, 21,* 153–174.

Mayes, B. T., & Ganster, D. C. (1983). A multitrait-multimethod matrix analysis of the PRF and MNQ need scales. *Journal of Management, 9,* 113–126.

Mellon, P. M., & Crano, W. D. (1977). An extension and application of the multitrait-multimethod matrix technique. *Journal of Educational Psychology, 69,* 716–723.

Murphy, K. R. (1982). Difficulties in the statistical control of halo. *Journal of Applied Psychology, 67,* 161–164.

Ostrom, T. M. (1969). The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Experimental Social Psychology, 5,* 12–30.

Ray, M. L., & Heeler, R. M. (1975). Analysis techniques for exploratory use of the multitrait-multimethod matrix. *Educational and Psychological Measurement, 35,* 255–265.

Roberts, M. A., Milich, R., Loney, J., & Caputo, J. (1981). A multitrait-multimethod analysis of variance of teachers' ratings of aggression, hyperactivity, and inattention. *Journal of Abnormal Child Psychology, 9,* 371–380.

Roshal, S. M., Frieze, I., & Wood, J. T. (1971). A multitrait-multimethod validation of measures of student attitudes toward school, toward learning, and toward technology in sixth grade children. *Educational and Psychological Measurement, 31,* 999–1006.

Schlesinger, I. M., & Guttman, L. (1969). Smallest space analysis of intelligence and achievement tests. *Psychological Bulletin, 71,* 95–100.

Schmitt, N. (1978). Path analysis of multitrait-multimethod matrices. *Applied Psychological Measurement, 2,* 157–178.

Schmitt, N., Coyle, B. W., & Saari, B. B. (1977). A review and critique of analyses of multitrait-multimethod matrices. *Multivariate Behavioral Research, 12,* 447–478.

Schmitt, N., & Saari, B. B. (1978). Behavior, situation, and rater variance in descriptions of leader behaviors. *Multivariate Behavioral Research, 13,* 483–495.

Schriesheim, C. A. (1981a). The effect of grouping or randomizing items on leniency response bias. *Educational and Psychological Measurement, 41,* 401–411.

Schriesheim, C. A. (1981b). Leniency effects on convergent and discriminant validity for grouped questionnaire items: A further investigation. *Educational and Psychological Measurement, 41,* 1093–1099.

Schriesheim, C. A., & DeNisi, A. S. (1980). Item presentation as an influence on questionnaire validity: A field experiment. *Educational and Psychological Measurement, 40,* 175–182.

Stanley, J. C. (1961). Analysis of unreplicated three-way classifications, with applications to rater bias and trait independence. *Psychometrika, 26,* 205–219.

Tinsley, H. E., & Kass, R. A. (1980). The construct validity of the leisure activities questionnaire and of the paragraphs about leisure. *Educational and Psychological Measurement, 40,* 219–226.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika, 31,* 279–311.

Tucker, L. R. (1967). Three-mode factor analysis of Parker-Fleishman complex tracking behavior data. *Multivariate Behavioral Research, 2,* 139–151.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1–10.

Turnage, J. J., & Muchinsky, P. M. (1982). Transsituational variability in human performance within assessment centers. *Organizational Behavior and Human Performance, 30,* 174–200.

Watkins, D., & Hattie, J. (1981). An investigation of the construct validity of three recently developed personality instruments: An application of confirmatory multimethod factor analysis. *Australian Journal of Psychology, 33,* 277–284.

Werts, C. E., & Linn, R. L. (1970). Path analysis: Psychological examples. *Psychological Bulletin, 74,* 193–212.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9,* 1–26.

Acknowledgements

The authors wish to acknowledge the helpful comments of two anonymous reviewers and the journal editor on several drafts of this paper.

Author's Address

Send requests for further information to Neal Schmitt, Department of Psychology, Michigan State University, East Lansing MI 48824-1117.

Reprints

Reprints of this article may be obtained *prepaid* for $2.50 (U.S. delivery) or $3.00 (outside U.S.; payment in U.S. funds drawn on a U.S. bank) from Applied Psychological Measurement, N658 Elliott Hall, University of Minnesota, Minneapolis, MN 55455, U.S.A.