DOCUMENT RESUME

ED 280 880 TM 870 210

AUTHOR Marsh, Herbert W.; Hocevar, Dennis
TITLE A New, More Powerful Approach to Multitrait-Multimethod Analyses: An Application of Second-Order Confirmatory Factor Analysis.
PUB DATE 8 Oct 86
NOTE 29p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Computer Software; Construct Validity; Correlation; *Error of Measurement; *Factor Analysis; Factor Structure; Goodness of Fit; *Mathematical Models; Measurement Techniques; *Multitrait Multimethod Techniques; Rating Scales; Research Problems; Test Items; *Test Theory
IDENTIFIERS *Confirmatory Factor Analysis; LISREL Computer Program; Students Evaluation of Educational Quality

ABSTRACT

The advantages of applying confirmatory factor analysis (CFA) to multitrait-multimethod (MTMM) data are widely recognized. However, because CFA as traditionally applied to MTMM data incorporates single indicators of each scale (i.e., each trait/method combination), important weaknesses are the failure to: (1) correct appropriately for measurement error in scale scores; (2) separate error due to low internal consistency from uniqueness due to weak trait and/or method effects; (3) test whether items or subscales accurately reflect the intended factor structure; and (4) test for correlated uniquenesses. However, when the analysis begins with multiple indicators of each scale (i.e., items or subscales), second-order factor analysis can be used to address each of these problems. In this approach first-order factors defined by multiple items or subscales are posited for each scale, and the method and trait factors are posited as second-order factors. This paper illustrates models that incorporate multiple indicators of each scale. The advantages of their application are discussed. A three-page list of references and the description of the models, mentioned in the text, supplement the paper. (Author/JAZ)

# A New, More Powerful Approach to Multitrait-Multimethod Analyses: An Application of Second-order Confirmatory Factor Analysis



Herbert W. Marsh
University of Sydney, Australia

Dennis Hocevar
University of Southern California

Revised: 24 June, 1986
Revised: 8 October, 1986

The authors would like to acknowledge Peter Bentler, Mark Groves, Bengt Muthen, and Jeffery Tanaka for their helpful comments on earlier drafts of this paper. Requests for reprints should be sent to Herbert W. Marsh, Department of Educational Psychology, University of Sydney, Sydney, NSW 2006 Australia.

Running Head: MTMM Analysis

A New, More Powerful Approach to Multitrait-Multimethod Analyses: An Application of Second-order Confirmatory Factor Analysis


ABSTRACT

The advantages of applying confirmatory factor analysis (CFA) to multitrait-multimethod (MTMM) data are widely recognized. However, because CFA as traditionally applied to MTMM data incorporates single indicators of each scale (i.e., each trait/method combination), important weaknesses are the failure to: a) correct appropriately for measurement error in scale scores; b) separate error due to low internal consistency from uniqueness due to weak trait and/or method effects; c) test whether items or subscales accurately reflect the intended factor structure; and d) test for correlated uniquenesses. However, when the analysis begins with multiple indicators of each scale (i.e., items or subscales), second-order factor analysis can be used to address each of these problems. In this approach first-order factors defined by multiple items or subscales are posited for each scale, and the method and trait factors are posited as second-order factors.

A New, More Powerful Approach to Multitrait-Multimethod Analyses: An Application of Second-order Confirmatory Factor Analysis

Campbell and Fiske (1959) advocate the assessment of construct validity by obtaining measures of more than one trait, each of which is assessed by more than one method. In analyzing multitrait-multimethod (MTMM) data it is typical to assess convergent validity, discriminant validity, and method/halo effects. Convergent validity is agreement between measures of the same trait assessed by different methods. Discriminant validity refers to the distinctiveness of the different traits. Method/halo effect is an undesirable bias that inflates the correlations among the different traits that are measured by the same method.

Determination of convergent and discriminant validity is based on inspection or analysis of a MTMM matrix. The original guidelines developed by Campbell and Fiske (1959) are still useful (Marsh & Hocevar, 1983). However, they have been criticized (Althauser & Heberlein, 1970; Alwin, 1974; Sullivan & Feldman, 1979; Widaman, 1985) particularly because they "are evaluated on the observed correlations among measures so that differences among variables in their level of reliability will distort both correlations among measures and summary measures derived from the correlations" (Widaman, 1985, p. 2). More recently other procedures have been developed for the analysis of MTMM matrices and confirmatory factor analysis (CFA) approaches have been widely recommended (e.g., Boruch & Wolins, 1970; Joreskog, 1974; Kenny, 1979; Forsythe, McGaghie & Friedman, 1986; Lomax & Algina, 1979; Marsh & Hocevar, 1983; 1984; Schmitt, 1978; Schmitt & Stults, 1986; Sullivan & Feldman, 1979; Werts, Joreskog & Linn, 1972; Werts & Linn, 1970; Widaman, 1985).

<u>The Traditional CFA Approach to MTMM Data.</u>

In the CFA of MTMM data factors defined by multiple indicators of the same trait support the construct validity of the trait, whereas factors defined by variables representing the same method argue for method/halo effects. In this approach a variety of CFA models with <u>a priori</u> factors corresponding to traits and/or methods are posited (see Marsh & Hocevar, 1983; Widaman, 1985). The parameter estimates and the ability of alternative models to fit the data are then used to assess convergent and discriminant validity, and method/halo effects. In the traditional CFA approach to MTMM data with LISREL-type analyses (e.g., Forsythe, et al., 1986; Marsh & Hocevar, 1983; Schmitt & Stults, 1986; Widaman, 1985), so long as there are at least three traits and three methods, the general CFA model is typically defined as follows:

1) Each scale (i.e., the nine or more distinct trait/method combinations) is inferred on the basis of only one measured variable, often an average of several items or subscales designed to measure that scale.

2) Factor loadings are constrained in LAMBDA Y such that each scale loads on one method factor and on one trait factor but all other factor loadings are fixed to be zero. One trait-factor is posited for each of the multiple traits and one method-factor is posited for each of the multiple methods.

3) Correlations among these factors in PSI are constrained such that correlations among method factors and among trait factors are freely estimated, but correlations between method and trait factors are fixed to be zero (see Joreskog, 1974; Marsh & Hocevar, 1983; Widaman, 1985).

4) The error/uniqueness of each scale is estimated in the diagonal of THETA and off-diagonals are fixed to be zero so that scale error/uniquenesses are uncorrelated.

Problems With The Traditional CFA Approach to MTMM Data.

Measurement error. The conceptualization of measurement error in this traditional CFA approach to MTMM data differs drastically from that of classical measurement theory. Because only one indicator of each scale is considered, a scale's reliability cannot be estimated on the basis of agreement among or the internal consistency of multiple indicators of the scale. Rather, measurement error is inferred from a scale's uniqueness -- its observed variation that cannot be explained by the other scales that are considered in the analysis. Whereas this estimate of measurement error contains random error variance as conceptualized in classical measurement theory, it also contains true score variance that is unique from the variance explained by the other scales (sometimes called specific variance). In contrast, measurement error conceptualized in classical measurement theory depends on the relative agreement among multiple indicators of the same scale and not the other scales that are included in the analysis. Consider, for example, a scale that is relatively unique from other scales in the analysis but is defined by items or subscales that are highly correlated. According to classical measurement theory this scale is very reliable because its multiple indicators are highly correlated. However, according to the traditional CFA approach to MTMM data this scale lacks reliability because it has a large component of unique or specific variance.

This inability to separate true uniqueness from random error is not limited to the CFA of MTMM data and Mulaik (1972, pp. 97-99) describes the same phenomena in terms of the general factor analysis model. Only when the items (or subscales) used to define each scale are included in the CFA will the conceptualization of measurement error in CFA will be like that of classical measurement theory. In this case the "uniqueness" of an item from other items designed to measure the same scale is legitimately considered to be error variance. Hence, both the CFA and classical measurement approaches to reliability infer measurement error from the lack of agreement among items designed to measure the same scale.

The conceptualization of measurement error in the traditional CFA approach to MTMM data creates important problems that may undermine its value. In particular, it is impossible to separate measurement error (random error variance) from uniqueness (specific variance) due to a lack of trait and method effects. Thus a low loading on a method or trait factor may be due to either substantial measurement error or a true lack of trait and method effects. The implications of this problem are particularly serious when the reliabilities of different scales vary as is typical in MTMM data. Such differences will distort inferred relations among the scales, the factor loadings on the latent method and trait factors, relations among the latent factors, and summary statistics that are based on these parameter estimates. Ironically, this is similar to the criticism of the Campbell-Fiske criteria and is often cited as an advantage of the CFA approach.

The construct validity of the scale scores. Multiple items designed to reflect each scale are typically averaged to form scale scores and MTMM analyses begin with these scale scores. Implicit in this process is the assumption that the researcher's a priori structure (i.e., the one implied by how scores are combined) accurately reflects the true factor structure. Unless there is empirical support for this a priori factor structure, however, the interpretation of the MTMM results may be problematic. If items from the same scale actually reflect different traits, or items from different scales actually reflect the same trait, then scale scores cannot be interpreted in terms of trait and method effects. The use of just a single item to represent each scale offers no solution to this problem, but merely precludes tests of the a priori factor structure. This problem is relevant to any approach to MTMM data that begins with scale scores, but it is ironic that traditional CFA approaches to MTMM analyses suffer from this problem even when multiple indicators of each scale are collected.

The hypothesized factor structure used to form scale scores is rarely tested in MTMM studies. In recognition of this problem, Marsh (1983) recommended that an exploratory factor analysis should conducted on item responses and that subsequent MTMM analyses should be based on the factors derived from this analysis instead of scale scores. If this recommendation were translated into the CFA approach it would involve first conducting a CFA on item or subscale scores. So long as there was reasonable support for the fit of the a priori structure, correlations among the factors derived from this CFA of item or subscale responses could then be used as the starting point for the traditional approach to the CFA of MTMM data. However, as described below, this actually corresponds to a second-order factor analysis in which the items or subscales are the measured variables, the factors from the first factor analysis disattenuated for measurement error are the first-order factors, and the factors from the second factor analysis are second-order factors used to infer trait and method effects.

The Application of HCFA to MTMM data in the Present Investigation.

CFA has been described in detail (see Bagozzi, 1980; Joreskog, 1981; Joreskog & Sorbom, 1983; Long, 1983; Pedhauzur, 1982) and has been frequently applied to MTMM data as described above. However, there have been few published applications of HCFA (Marsh, 1985; in press-a; Marsh & Hocevar, 1985), and we know of no previous applications of HCFA to MTMM data. Hence, the purpose of the present investigation is to describe this new approach to MTMM data that incorporates multiple indicators of each scale and to demonstrate its advantages over the traditional CFA approach.

Conceptually, hierarchical factor analysis would be like conducting a factor analysis on a correlation matrix of measured variables, estimating correlations among the first-order factors, and then doing a second factor analysis on the correlations among the first-order factors. The results of this second factor analysis are used to infer second-order factors that are derived from relations among the first-order factors. In the HCFA approach, however, both first-order and second-order factors are estimated simultaneously in the same analysis.

There are alternative parameterizations of the HCFA model (e.g., Bentler & Weeks, 1980; Joreskog, 1974; Joreskog & Sorbom, 1981; 1983; Marsh & Hocevar, 1985; McDonald, 1985; Olson, 1982). We used Joreskog and Sorbom's submodel 3B (Joreskog & Sorbom, 1983, pp. I.11; also see Marsh and Hocevar, 1985; and Olson; 1982) and matrix definitions from LISREL VI. For this parameterization the HCFA model is defined in terms of four parameter matrices: LAMBDA Y (LAMBDA), the matrix of first-order factor loadings; BETA, the matrix of second-order factor loadings; PSI, the matrix of residual factor variances for first-order factors (and correlations among residuals if posited) and factor variances and covariances for second-order factors; and THETA EPSILON (THETA), the matrix of error/uniquenesses of the measured variables (and correlations among uniquenesses if posited). Using this specification and I=the identity matrix, the covariance matrix (S) of the observed y-variables is:

(1) $S = \text{LAMBDA} \ (I - \text{BETA})^{-1} \ \text{PSI} \ (I - \text{BETA})^{-1\prime} \ \text{LAMBDA}^{\prime} + \text{THETA}$

If second order factors are not hypothesized, then BETA=0 and this model becomes the traditional first-order factor analysis model:

(2) $S = \text{LAMBDA} \ \text{PSI} \ \text{LAMBDA}^{\prime} + \text{THETA}$

For purposes of the present investigation there are 27 measured variables (item responses), three measured variables define each of 9 scales, and each scale represents a unique combination of one of 3 traits and one of 3 methods (see Figure 1). In the traditional CFA approach, responses to the three items for each scale would be combined according to an a priori, untested hypothesis. The CFA would then be applied to the uncorrected 9 x 9 correlation matrix of scale scores. In the HCFA approach, correlations among the 27 measured variables are used to define nine first-order factors. These nine factors form the basis of subsequent analyses instead of the nine scale scores used in the CFA approach. Covariances among these first-order factors are used to define second-order factors that represent trait and method variance (Figure 1). As will be shown, this HCFA approach corrects the MTMM matrix for unreliability whereas the CFA does not and it provides a test of the validity of the hypothesized first-order factor structure whereas the CFA approach does not.

## Method

### The Data.

The data represent students' evaluations of teaching effectiveness as measured by the Students' Evaluations of Educational Quality (SEEQ) instrument. SEEQ consists of 35 items designed to measure nine factors that have been identified in numerous factor analytic studies (e.g., Marsh, 1984). For purposes of this demonstration 3 items designed to measure each of 3 factors (Learning/Value, Group Interaction, and Workload/Difficulty), a total of 9 items, were selected. Class-average ratings for 948 different classes were selected such that there were three sets of ratings for each of 316 instructors teaching the same class on three different occasions (see Marsh & Hocevar, 1984, for further description). The 3 evaluation factors represent the multiple traits, the 3 occasions represent the multiple methods, and each of the 9 method/trait combinations is measured by 3 items. Hence the MTMM matrix consists of correlations among 27 measured variables

(3 items per trait x 3 traits x 3 occasions) for each of 316 instructors. Researchers have frequently considered different occasions as the multiple methods in MTMM studies (e.g., Campbell & O'Connell, 1967; Sullivan & Feldman, 1979; Werts, Joreskog & Linn, 1972). In the present investigation the same instructor was evaluated on each occasion, but the ratings were completed by different groups of students. Hence, ratings collected on occasion 1 have in common only the fact that the instructors being evaluated have taught the course less frequently than for ratings collected on occasion 3. Thus the method effects for this MTMM study are likely to be small.

Statistical Analyses.

The commercially available LISREL VI program (Joreskog & Sorbom, 1983) was used for all statistical analyses and was the basis for the notation and specification of models. LISREL, after testing for identification, attempts to reproduce the observed correlation matrix under the constraints of a hypothesized model. A $X^2$ test is used to assess whether or not residual differences between the observed and hypothesized covariance matrices will converge to zero as the sample size tends to infinity. However, hypothesized models are best regarded as approximations to reality rather than exact statements of truth so that any model will be rejected if the sample size is large enough (Cudeck & Browne, 1983). As noted by Marsh and Hocevar (1985, p. 567), "most applications of confirmatory factor analysis require a subjective evaluation of whether or not a statistically significant chi-square is small enough to constitute an adequate fit." The problem of goodness of fit is how to decide whether the residuals are sufficiently small to justify the conclusion that a specific model adequately fits the data. Many alternative indices of fit have been developed including: a) the $X^2$/df ratio; b) the root mean square residual (RMSR) based on differences between the original and reproduced correlation matrices; and c) the Tucker-Lewis index and the Bentler-Bonett index (BBI) that provide an indication of the proportion of variance that is explained by the hypothesized model. Though none of these alternative indices has been universally endorsed, we use each in order to assess goodness of fit (see Bentler & Bonett, 1980; Cudeck & Browne, 1983; Long, 1983; Marsh, Balla, McDonald, 1986; Marsh & Hocevar, 1985, for further discussion).

RESULTS

The First-Order Model.

The first-order model as a target model. The $X^2$ for a HCFA model can be only as good as the fit of the corresponding first-order model. In this respect the first-order model represents an upper-bound or optimum for all subsequent HCFA models, and so we refer to the first-order model (Model 1) as a target model (see Marsh & Hocevar, 1985). The fit of this first-order model is also important because unless the parameter estimates support the a priori factor structure and the fit is reasonable, then subsequent interpretations of trait and method factors may be unjustified. The pattern of factor loadings used to define the first-order factors is shown in Table 1. Each of the nine first-order factors represents a unique trait-method combination that is defined by three items. One item for each factor is designated to be a reference indicator and its factor loading is fixed to be 1.0. Inspection of the factor loadings for the target model (not shown since these are nearly the same as those in Table 1) indicated that each of the 9 first-order factors was well defined in that: a) all factor loadings and factor variances differed from zero by at least 15 standard errors; b) standardized factor loadings (not shown) were generally .9 or higher; and c) LISREL VI also provides a 27x9 matrix of correlations between the 27 measured variables and the 9 first-order factors, and inspection of this matrix indicated that every item was substantially correlated with its posited factor and substantially less correlated with each of the other factors (see Hocevar & El-Zahhar, 1985, for further discussion of this approach).

The MTMM matrix based on the first-order model. When a reference indicator is used to determine the metric of a factor, the factor variances in the diagonal of PSI are freely estimated and factor covariances appear in the off-diagonals of PSI. In standardized form (Table 2) this is a correlation matrix of relations among the first-order latent factors. Superficially, this 9 x 9 correlation matrix of relations among latent constructs is like MTMM matrices based on correlations among scale scores such as used in the traditional CFA approach. However, the correlations in Table 2 are based on an optimally weighted combination of the multiple indicators and are corrected for measurement error (see discussion below). Hence, the Campbell-Fiske criteria can be more appropriately applied to the correlations in Table 2 than to correlations among scale scores. Inspection of these correlations suggests that support for convergent and discriminant validity is strong whereas support for method effects is weak. A more precise quantification of these effects requires the application of HCFA.

Correction for measurement error. In the traditional approach to CFA of MTMM data correlations are based on observed scale scores even when these measured variables are the mean of multiple indicators. Thus, the scale scores and the correlations among them reflect an unspecified amount of measurement error. In LISREL, as in classical measurement theory, the correlation between two scale scores corrected for attenuation due to error variance is the correlation between their true scores. The correlation between true scores is estimated by the correlation between the measured variables divided by the geometric mean of appropriately determined reliabilities of the measures. In LISREL, however, error variance is estimated directly in THETA (see equation 2) instead of being inferred on the basis of reliability estimates. The relation between this correction for attenuation and the corrections for unreliability using the traditional formulae depends on the reliability estimate used and the appropriateness of the reliability estimates. It should be noted, however, that this correction for unreliability is based on the first-order model; it is not a function of the HCFA model, but is the normal correction for unreliability incorporated into the LISREL analyses.

The reliability of a linear combination of multiple indicatators is the ratio of true score variance to total observed variance. Cronbach's alpha, as typically applied, provides an estimate of this reliability when measures are parallel (i.e., tau-equivalent). Because LISREL uses an optimal weighting of the multiple indicators Cronbach's alpha underestimates reliability unless all the factor loadings are equal (Kenny, 1979; McDonald, 1985). Kenny (1979) demonstrated how reliability estimates can be derived from the maximum likelihood factor loadings and correlations among the multiple indicators, and how these differ from the reliability of other linear composites. Similarly, McDonald (1985) demonstrated how the reliability can be estimated from just the factor loadings and the error/uniqueness terms (in the diagonal of THETA). Hence, the correction for attenuation in LISREL and more traditional approaches differ according to the appropriateness of the reliability estimates used rather than the operationalization of this correction once an appropriate estimate of reliability has been established. As noted by McDonald (1985) the advantage of the factor analytic perspective to reliability theory is that it provides information about about the characteristics of the items or subscales used to estimate a construct (e.g., whether they are parallel) as well as the information needed to estimate reliability (see Joreskog, 1981, for further discussion of the relation between classical test theory and LISREL).

The validity of the a priori factor structure used to combine items.

The traditional approach to CFA of MTMM data begins with correlations based on observed scale scores so that the measured variables have an unknown relation to the latent constructs that they are designed to measure. There is typically no attempt to test the validity of the *a priori* factor structure implicit in the way multiple indicators are combined. In the HCFA approach the goodness of fit and parameter estimates for the first-order model provide a precise indication of how well the latent constructs are defined by the items or subscales used to infer the latent constructs. Furthermore, the latent construct is based on an optimal weighting of the multiple indicators. An unweighted sum of the multiple indicators of each scale may provide an adequate representation of the data, but the HCFA approach provides an empirical test of this implicit assumption.

*HCFA Models of the MTMM data.*

In the general HCFA model (Model 2), the pattern of the first-order factor loadings is the same as for the target model (Model 1) and first-order loadings (Table 1) were virtually the same as those from Model 1. However, the 36 covariances among the first-order factors (in PSI) were all set to be zero and six additional second-order factors were posited to represent the 3 traits and 3 methods (see Figure 1). The covariances among first-order factors are set to zero because the purpose of the second-order factors is to provide an alternative explanation of these covariances. In the HCFA model the diagonal elements in PSI associated with these first-order factors are residual variances. Because the measurement error variances are contained in the diagonal of THETA, the residual variances in PSI represent uniquenesses -- variance in the first-order factors that cannot be explained by second-order factors. In the LISREL specification used here, the factor loadings for the second-order factors are estimated in BETA, second-order factor variances are estimated in PSI, and covariances among the second-order trait factors and among the second-order method factors are estimated in PSI[1]. Each of these 6 second-order factors is defined by 3 first-order factors, and one of these first-order factors is designated to be a reference indicator whose factor loading is set to be 1.0.

The second-order factor loadings (Table 3) for the method factors are modest and several are not statistically significant. The loadings for the trait factors are much larger and each loading is between 8 to 20 times the size of its standard error. Similarly, the variances of method factors are small or nonsignificant relative to their standard errors, whereas variances of the trait-factors are all significant and much larger. Collectively, The traditional approach to CFA of MTMM data begins with correlations based on observed scale scores so that the measured variables have an unknown relation to the latent constructs that they are designed to measure. There is typically no attempt to test the validity of the a priori factor structure implicit in the way multiple indicators are combined. In the HCFA approach the goodness of fit and parameter estimates for the first-order model provide a precise indication of how well the latent constructs are defined by the items or subscales used to infer the latent constructs. Furthermore, the latent construct is based on an optimal weighting of the multiple indicators. An unweighted sum of the multiple indicators of each scale may provide an adequate representation of the data, but the HCFA approach provides an empirical test of this implicit assumption.

HCFA Models of the MTMM data.

In the general HCFA model (Model 2), the pattern of the first-order factor loadings is the same as for the target model (Model 1) and first-order loadings (Table 1) were virtually the same as those from Model 1. However, the 36 covariances among the first-order factors (in PSI) were all set to be zero and six additional second-order factors were posited to represent the 3 traits and 3 methods (see Figure 1). The covariances among first-order factors are set to zero because the purpose of the second-order factors is to provide an alternative explanation of these covariances. In the HCFA model the diagonal elements in PSI associated with these first-order factors are residual variances. Because the measurement error variances are contained in the diagonal of THETA, the residual variances in PSI represent uniquenesses -- variance in the first-order factors that cannot be explained by second-order factors. In the LISREL specification used here, the factor loadings for the second-order factors are estimated in BETA, second-order factor variances are estimated in PSI, and covariances among the second-order trait factors and among the second-order method factors are estimated in PSI[1]. Each of these 6 second-order factors is defined by 3 first-order factors, and one of these first-order factors is designated to be a reference indicator whose factor loading is set to be 1.0.

The second-order factor loadings (Table 3) for the method factors are modest and several are not statistically significant. The loadings for the trait factors are much larger and each loading is between 8 to 20 times the size of its standard error. Similarly, the variances of method factors are small or nonsignificant relative to their standard errors, whereas variances of the trait-factors are all significant and much larger. Collectively, these findings indicate traits effects (i.e., convergent validities) are much larger than method/halo effects.

Covariances among method factors (Table 3) are all nonsignificant, whereas 2 of 3 covariances among trait factors are significant. Learning/Value correlates (i.e., the standardized covariance) .55 with Group Interaction and .34 with Workload/Difficulty. These correlations represent estimated relations among latent true scores that have been corrected for measurement error as described earlier.[2] In the terminology of MTMM analyses these represent true trait-covariances.

Trait-only (Model 3) and method-only (Model 4) models were also defined by including only three second-order factors (i.e, 3 trait factors in the trait-only model and 3 method factors in the method-only factor). In each case the restricted model provided a poorer fit to the data (Table 4) than Model 2. However, the trait-only model provides a substantially better fit than does the method-only model. These results also demonstrate that trait factors account for much more variance than method factors.

In summary, the results demonstrate: a) a well defined factor structure as indicated by the high first-order loadings; b) weak method effects as indicated by the generally low loadings on second-order method factors, by the small variances of these second-order factors, and by the small decrement in fit produced by eliminating the method factors in Model 3; c) strong convergent validity as indicated by the high loadings on the second-order trait factors, by the substantial variances of these second-order factors, and by the substantial decrement in fit produced by eliminating the trait-factors in Model 4; and d) low to moderately correlated true trait-scores (see footnote 2).

<u>The Second-order CFA Model With Correlated Uniquenesses: An Expanded Formulation of Measurement Error.</u>

The fit of Model 2, the general HCFA structure, can be evaluated from different perspectives. It is well-defined in that: a) all first-order factor loadings and at least the second-order trait factor loadings are substantial; b) the BBI (.85) is reasonably high; and c) Model 2 fits the data nearly as well as its target model ($X^2_{diff}$ = 22, df=12). Nevertheless, the overall fit to the data may be less than satisfactory in that $X^2$/df ratio is 5.28. Problems inherent in the assessment of fit complicate interpretations of fit. Nevertheless, the comparison of Models 1 and 2 demonstrates that the lack of fit in Model 2 occurs in the definition of the first-order factors and not the second-order factors. That is, the fit of

Model 2 cannot possibly be better than Model 1 (i.e., the models are nested) and Model 2 fits the data almost as well as Model 1.

One possible reason for the relatively poor fit of Model 1 is that the simple structure that requires each measured variable to load on one and only one first-order factor is quite restrictive, and a better fit might result if measured variables were allowed to load on more than one first-order factor. However, such a model might also confound method and trait factors. Furthermore, the a priori nature of the data for this particular demonstration provides a more likely explanation. Models 1-4 all require the error/uniquenesses to be uncorrelated (i.e., THETA is a diagonal matrix). This assumption is common in psychological measurement and may be reasonable in many instances, but it may not be reasonable when the same item is used on different occasions (e.g., Feldman & Sullivan, 1979; Joreskog, 1974; 1979). For example, the same items designed to measure Learning/Value were administered on each occasion. As shown in Figure 2, the uniqueness of item 1 on occasion 1 is posited to be correlated with the uniqueness of item 1 on occasion 2. These correlated uniquenesses may effect the second-order trait or method factors as well as goodness of fit. It is also important to emphasize that this type of effect cannot typically be examined with the traditional CFA approach to MTMM data since only one indicator of each trait/method combination is considered.

In order to test this suggestion four new models were posited to have correlated uniquenesses. Each of these new models (Models 1a-4a) differed from the corresponding models described earlier (Models 1-4) only in that 27 off-diagonal elements in THETA EPSILON were set free. The elements were freed such that the uniqueness of each item was allowed to covary with the uniqueness of the same item administered on different occasions. A comparison of the fit indices for Models 1 and 1a (see Table 4) provides a test of the assumption of uncorrelated uniquenesses. Model 1a provides a much better fit to the data than Model 1 and results in substantial improvements in the other indices of fit. Inspection of THETA (not shown) indicated that 22 of the 27 posited correlations were statistically significant and all these were positive. Similarly, for each of the other pairs of models (Models 2 vs. 2a, 3 vs. 3a, and 4 vs.4a), the model with correlated uniquenesses fits the data substantially better. These results demonstrate that the assumption of uncorrelated uniquenesses is unjustified for these data.

Model 2a, the full HCFA model with correlated uniquenesses, fits the data nearly as well as the new target Model 1a. Thus the correlations among the first-order factors can still be adequately explained in terms of the second-order trait and method factors. Despite the substantial improvement in the overall fit produced by positing correlated uniquenesses, the factor loadings for the first- and second-order factors and correlations among the second order factors are nearly the same for Models 2 and 2a. Furthermore, the difference in $X^2$s for Models 1a and 2a is nearly the same as the difference between Models 1 and 2.

Trait-only (Model 3a) and method-only (Model 4a) with correlated uniquenesses were also fit to the data. These models again result in a poorer fit than the new target Model 1a, and the trait-only model again fits the data better than the method-only model. Furthermore, the differences in $X^2$s between these restricted models and the full HCFA model (i.e., 2a vs. 3a and 4a) are nearly the same as the corresponding differences for models with no correlated uniquenesses (see Table 4).

The correlated uniqueness associated with the idiosyncratic wording of specific items represents a second source of method effect that can be evaluated from the MTMM perspective. In fact, inspection of the fit indices in Table 4 indicates that the method effects associated with specific items accounted for much more variance than the method effects associated with with the multiple occasions of data collection. Method effects are typically represented as method factors in CFA models. However, Kenny (1979) proposed an alternative specification of the MTMM model in which method effects are represented as correlated uniqueneesses in THETA such as posited here and Marsh (in press-b) noted advantages of this representation. These results also demonstrate that more than one source of method effect can be considered in the same MTMM study. Elsewhere, though not based on HCFA models of MTMM data, I argued that MTMM studies that consider more than one source of method effect provide a stronger evaluation of construct validity than studies that consider only one source of method effect (Marsh, Barnes & Hocevar, 1985; Marsh & Butler, 1984).

In summary, each of the models positi j correlated uniquenesses (Models 1a-4a) provides a substantial improvement over the corresponding models without correlated uniquenesses (Models 1-4). This demonstrate that method effects associated with the idiosyncratic wording of specific items items did influence the ratings even though the ratings were made by different groups of students. Despite this improved fit, parameter estimates for the second-order trait factors and method factors associated with the multiple occasions were nearly unchanged. At least in this application, the improved fit is primarily a function of improved fit of the first-order factors and has nearly no impact on the second-order factor structure. Further research is needed to establish whether this finding is typical, though we suspect that the influence of correlated uniquenesses may be more substantial when the same subjects respond to the same materials on different occasions as in the typical panel design.

## Discussion and Implications

### A Comparison of the CFA and HCFA Approaches To MTMM Data.

The purpose of this demonstration is to describe new, more powerful models based on HCFA for the analysis of MTMM data. CFA approaches to MTMM data are used frequently but these approaches begin with a single indicator of each scale. The failure to incorporate multiple measures of each scale, even when available, constitutes a serious weakness in the traditional CFA approach to MTMM data. It is ironic that MTMM studies with their emphasis on multiple indicators, and particularly CFA approaches to MTMM data, have not incorporated information from the multiple indicators used to represent each trait/method combination. The HCFA approach to MTMM data described here differs from the typical CFA approach in three important ways.

First, estimates of measurement error in the inferred scale scores are based on the agreement among multiple indicators that measure the same latent construct in the HCFA approach rather than residual variance that is unexplained by other scales (uniqueness) as in the traditional CFA approach. In the HCFA approach random error inferred from low correlations among multiple indicators of the same scale is clearly separated from uniqueness due to weak trait and method effects, but in the traditional CFA approach the two sources are confounded. The definition of and correction for measurement error in the HCFA approach is clearly more consistent with traditional conceptualizations of classical measurement theory, and seems less arbitrary in that reliability estimates do not depend on what other scales are included in the analysis. In fact, the only justification for the conceptualization of and correction for measurement error in the traditional CFA approach seems to be its inability to separate measurement error from uniqueness.

A second difference between the two approaches is that the HCFA approach provides rigorous tests of the *a priori* factor structure posited to underlie the multiple indicators of each scale whereas the traditional CFA approach provides none. Typically, investigators merely sum responses to the items or subscales designed to measure each scale without testing their *a priori* structure (i.e., the one implied by the way they combine scores). In this traditional approach, the interpretation of trait- and method-factors will be problematic if the *a priori* structure does not fit the data, but tests of this fit are not considered. Even when the pattern of factor loadings is consistent with the *a priori* structure, the method of combining the measured variables to form scale scores (e.g., an unweighted sum) may be inconsistent. In the HCFA approach, parameter estimates and fit indices for the target model provide an empirical test of the *a priori* model and optimally defined factors. Furthermore, if the target model does not provide an adequate fit, it can be modified according to the substantive nature of the data or empirical guidelines. Higher-order method and trait factors can then be posited on the basis of the new first-order structure as in the present demonstration.

A third difference between the two approaches is that whereas both approaches typically posit uncorrelated error/uniquenesses for the items used to define each scale, this assumption is easily tested and modified with the HCFA approach but not with the CFA approach. In some MTMM applications, as in the present demonstration, the *a priori* nature of the data make this assumption problematic. The findings demonstrated that this assumption was unjustified, even though the correction for these correlated uniquenesses had nearly no effect on the substantive findings. However, it is possible that in other applications the inclusion of correlated errors will have an even larger effect on goodness of fit and also affect substantive conclusions -- particularly when the same subjects respond to the same materials on multiple occasions as in the typical panel design.

The HCFA approach to MTMM data has apparently not been previously considered, but the logic on which it is based is not new. Marsh (1983) recognized that scale scores are based on an *a priori* factor structure and he proposed the use of exploratory factor analyses to test this structure. Campbell and Fiske (1959, p. 102) noted that MTMM matrices must be evaluated in relation to the reliabilities of the scale scores but proposed no systematic approach to accomplish this. Althauser and Heberlein (1970) argued that under certain circumstances (e.g., measurement errors are uncorrelated) the entire MTMM matrix may be corrected for unreliability using conventional correction formulas. If the *a priori* factor structure implicit in the way subscale or item scores are combined matches the observed factor structure and if correlations among scale scores are corrected for measurement error inferred on the basis of agreement among multiple indicators, then the traditional CFA approach starting with this disattenuated MTMM matrix will be equivalent to the HCFA approach. Even in this ideal situation, however, the HCFA approach provides systematic tests underlying assumptions that are typically untestable with the CFA approach.

<u>The Generality and Potential Weaknesses in the HCFA Approach.</u>

In the present demonstration, the multiple traits were defined by the same set of items administered on different occasions, but the HCFA approach is more general and can be used in most applications that are appropriate for the traditional CFA approach. First, because it is not necessary that the different methods consist of different occasions, the HCFA approach can be used with all types of method variation. Second, it is not necessary to use the same items for defining the same trait with different methods so long as each trait/method combination is assessed with multiple indicators. In fact, the use of the same items for different methods in the present demonstration appears to be a major reason why the second-order factor model with uncorrelated uniquenesses (Model 2) failed to fit adequately the data. Third, even if some of the trait/method combinations are inferred with only a single indicator, it may still be possible to use the HCFA approach. Even though the single-item factors do not allow for estimates of internal consistency, reliability estimates can be incorporated into the model. Testing a plausible range of such estimates could be used to assess the sensitivity of the other parameter estimates to differences in these assigned reliability estimates (see Newman, 1984; Land & Felson, 1978). Fourth, whereas both the CFA and HCFA approaches normally require at least 3 traits and 3 methods, special CFA models have been developed for applications with only 2 traits (Kenny, 1979) or two methods (Marsh & Hocevar, 1983) and can be adapted to the HCFA approach. Furthermore, the variety of of alternative models considered in the CFA approach to MTMM data (e.g., Marsh, in press-b; Widaman, 1985) can be easily adapted to the HCFA approach (e.g., Models 3 and 4). Finally, even in MTMM studies in which each trait/method combination is inferred with a single indicator, the HCFA approach can be used. The HCFA approach would still have the benefit of separating error due to separate random error from uniqueness due to weak trait and/or method effects, though this might be of no advantage unless reasonable estimates of reliability were available. However, in some such applications where reliability estimates are available (e.g., published studies that provide reliability estimates) the HCFA approach would be clearly preferable. In summary, the HCFA approach to MTMM data has a wide range of application, can be used in most applications in which the traditional CFA approach is appropriate, and is recommended instead of the

CFA approach whenever multiple indicators of each trait/method combination are available or when the reliability these scales can be estimated.

An anonymous reviewer suggested several potential limitations of the HCFA approach to item level data that deserve further attention: a) items may be dichotomous and LISREL-type CFAs, like most other factor analytic techniques, assume continuous variables; b) item level data are typically less reliable --particularly when based on responses by individuals instead of groups -- so that they may be more prone to unstable or improper solutions (e.g., Heywood cases); c) psychological inventories often have many more items than in the present application so that costs of the HCFA approach may be prohibitive. In reference to the first point, Muthen has developed an appropriate mathematical solution for analyses of dichotomous variables and shown that LISREL-type estimates are quite robust though $X^2$ tests may be biased if item skews are extreme (Muethen & Kaplan, 1985). LISREL VI can analyze dichotomous variables though the corresponding $X^2$ test of significance must be ignored. The second point is intuitively reasonable, but further research is needed to test the suggestion. In fact, the traditional CFA approach is known to suffer from this problem and this may be related to the its inappropriate definition of and correction for measurement error. Thus it is possible that item level analyses will be <u>less</u> prone to improper solutions. The third point suggests a practical limit to the number of items that can be factor analyzed. Such a limit will be a complicated function of the characteristics of the computer software, the computer itself, the model to be tested, and the congruence between the model and the data. A partial solution to all three potential limitations would be to divide all the items from each scale into three or more subscales to be used in further analyses. For example, Marsh and Hocevar (1985) conducted a HCFA on a 56 item self-concept inventory by forming 28 two-item subscales that were used to define 7 first-order factors. This procedure, while sacrificing some item-level information, preserves most of the advantages of the HCFA approach described here, eliminates dichotomous responses if they exist, increases the reliability of the measured variables, and reduces the number of measured variables to 3 or 4 times the number of first-order factors.

## Footnotes

[1] In the LISREL specification of Models 2 and 2a, 21 measured variables were used to represent 15 factors -- 9 first-order and 6 second-order factors (see Figure 1): Lambda Y was a 27x15 matrix with the first 9 columns structured as in Table 2 and the last 6 columns as fixed zeros; Beta was a 15x15 matrix with fixed zeros in the first 9 columns and the last 6 columns structured as in Table 3; PSI was a 15x15 matrix with fixed zeros in the non-diagonals values of the first 9 columns and the last 6 columns structured as in Table 3; Theta Epsilon was a 27x27 matrix that was diagonal for Model 2 but had 27 off-diagonals that were not fixed in Model 2a (see Figure 2).

[2] There is no universally accepted criteria of what constitutes support for discriminant validity in CFA approaches to MTMM data. Some researchers (e.g., Kenny, 1979; Lomax & Algina, 1979; Widaman, 1985) argue that correlations between trait factors provide evidence against discriminant validity. In contrast, Werts, Linn and Joreskog (1971) note that there are no rules for determining how high true-trait correlations should be before before the traits are considered to be indistinguishable other than to test whether the correlation differs significantly from unity. Similarly, Marsh and Hocevar (1983) argue that significant correlations between trait-factors imply true trait correlations, and should only be interpreted as a lack of discriminant validity when such correlations approach unity or are inconsistent with the substantive nature of the data. In support of their argument they note that Campbell and Fiske's fourth guideline interprets the consistency of correlations among different traits across methods -- an indication of true trait-covariance -- as support for discriminant validity rather than evidence against it. This ambiguity, though an important issue, has not been emphasized in the present demonstration because it applies to both CFA and HCFA approaches.

REFERENCES

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

Alwin, D. F. (1974). Approaches to the interpretation of relationships and the multitrait-multimethod matrix. In H. L. Costner (Ed.), Sociological methodology 1973-4 (pp. 79-105). San Francisco: Jossey-Bass.

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

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

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

Forsythe, G. B., McGaghie, W. B., & Friedman, C. P. (1986). Construct validity of medical clinical competence measures: A multitrait-multimethod study using confirmatory factor analysis. American Educational Research Journal, 23, 315-336.

Hocevar, D. & El-Zahhar, N. E. (1985). Test anxiety in the USA and Egypt: A paradigm for investigating psychometric characteristics across cultures. In. H. M. Van Der Ploeg, R. Schwarzer, & C. D. Spielberger (Eds.), Advances in test anxiety research (Vol. 4, pp. 203-213). Swets & Zeitlinger.

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

Joreskog, K. G. (1979). Statistical estimation of structural models in longitudinal investigations. In J. R. Nesselroade & B. Baltes (Eds.), Longitudinal research in the study of behavior and development (pp. 303-351). New York: Academic Press.

Joreskog, K. G. (1980). Structural analysis of covariance and correlation matrices. Psychometrika, 43, 443-477.

Joreskog, K. G. (1981). Analysis of covariance structures. Scandanavian Journal of Statistics, 8, 65-92.

Joreskog, K. G. & Sorbom, D. (1981). LISREL V: Analysis of Linear Structural Relations By the Method of Maximum Likelihood. Chicago: International Educational Services.

Joreskog, K. G. & Sorbom, D. (1983). LISREL VI: Analysis of Linear Structural Relations By the Method of Maximum Likelihood. Chicago: International Educational Services.

Kenny, D. A. (1979). Correlation and causality. New York: Wiley.

Land, K. C., & Felson, M. (1978). Sensitivity analysis of arbitrarily identified simultaneous-equation models. Sociological Methods and Research, 6, 283-307.

Lomax, R. G., & Algina, J. (1979). Comparison of two procedures for analyzing multitrait-multimethod matrices. Journal of Educational Measurement, 16, 177-186.

Long, K. S. (1983) Confirmatory factor analysis: A preface to LISREL. Beverly Hills, CA: Sage.

Marsh, H. W. (1983). Multitrait-multimethod analysis: Distinguishing between items and traits. Educational and Psychological Measurement, 43, 231-248.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. Journal of Educational Psychology, 76, 707-754.

Marsh, H. W. (1985). The structure of masculinity/femininity: An application of confirmatory factor analysis to higher-order factor structures and factorial invariance. Multivariate Behavioral Research, 20, 427-449.

Marsh, H. W. (in press-a). The hierarchical structure of self-concept: An application of hierarchical confirmatory factor analysis. Journal of Educational Measurement.

Marsh, H. W. (in press-b). Multitrait-multimethod analyses. In J. P. Keeves (Ed.), The international handbook of educational research methodology, measurement and evaluaion. New York: Pergamon Press.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1986). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. (In Review).

Marsh, H. W., Barnes, J., & Hocevar, D. (1985). Self-other agreement on multidimensional self-concept ratings: Factor analysis & multitrait-multimethod analysis. Journal of Personality and Social Psychology, 49, 1360-1377.

Marsh, H. W. & Butler, S. (1984). Evaluating reading diagnostic tests: An application of confirmatory factor analysis to multitrait-multimethod data. Applied Psychological Measurement, 8, 307-320.

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. (1984). The factorial invariance of students' evaluations of college teaching. American Educational Research Journal, 21, 341-366.

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

Muthen, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrica, 49, 115-132.

Muthen, B. & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal likert variables. British Journal of Mathematical and Statistical Psychology, 38, 171-189.

Newman, R. S. (1984). Children's achievement and self-evaluations in mathematics: A longitudinal study. Journal of Educational Psychology, 76, 857-873.

Olson, G. H. (1982, March). Covariance structure analysis involving measured variables assumed to have higher-ordered factor structures. Paper presented at the American Educational Research Association, Montreal.

Pedhauzur, E. J. (1982). Multiple regression in behavioral research (2nd ed.). New York: Holt, Rinehart and Winston.

Schmitt, N. (1978). Path analysis of multitrait-multimethod matrices. Applied Psychological Measurement, 2, 157-173.

Schmitt, N., & Stults, D. M. (1986). Methodological review: Analysis of multitrait-multimethod matrices. Applied Psychological Measurement, 10, 1-22.

Sullivan, J. L., & Feldman, S. (1979). Multiple indicators: An introduction. Beverly Hills, CA: Sage.

Werts C. E., Joreskog, K. G., & Linn, R. L. (1972). A multitrait-multimethod model for studying growth. Educational and Psychological Measurement, 32, 655-678.

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.

Figure 1. The structure of measured variables, first-order factors, second-order method (M) factors and second-order trait (T) factors. Each of the 27 measured variables (item responses) has a three-part label corresponding to the item (1-9), the method (M1-M3) of assessment, and the trait (T1-T3) that the item is designed to measure. Each of the first-order factors represents a distinct trait/method combination (M1T1-M3T3) that is inferred from responses to 3 measured variables. These first-order factors are used to define the 3 second-order method and trait factors. Although not shown in the figure, the three second-order method factors were assumed to be correlated as were the three second-order trait factors.

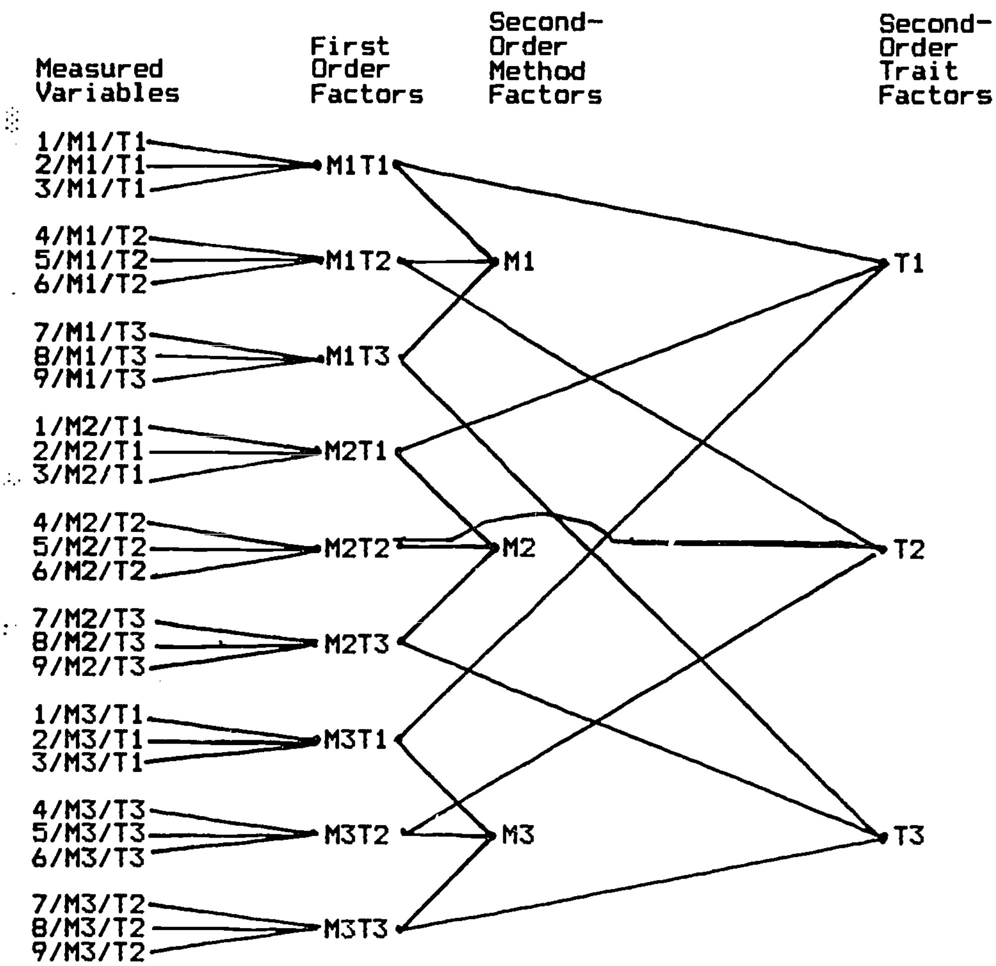

Figure 2. A Representation of Correlated Uniquenesses. T1 is the first second-order trait-factor, M1 and M2 are the first two second-order method-factors, M1T1 and M2T1 are first-order factors representing T1 assessed by the first two methods, 1/M1/T1, 2/M1/T1, 1/M2/T1 and 2/M2/T1 are measured variables used to infer the the first-order factors (see Figure 1) and the U associated with each measured variable represents its error/uniqueness. The curved arrows between uniquenesses represent the correlated uniquenesses of the same measured variables administered according to two different methods (occasions).

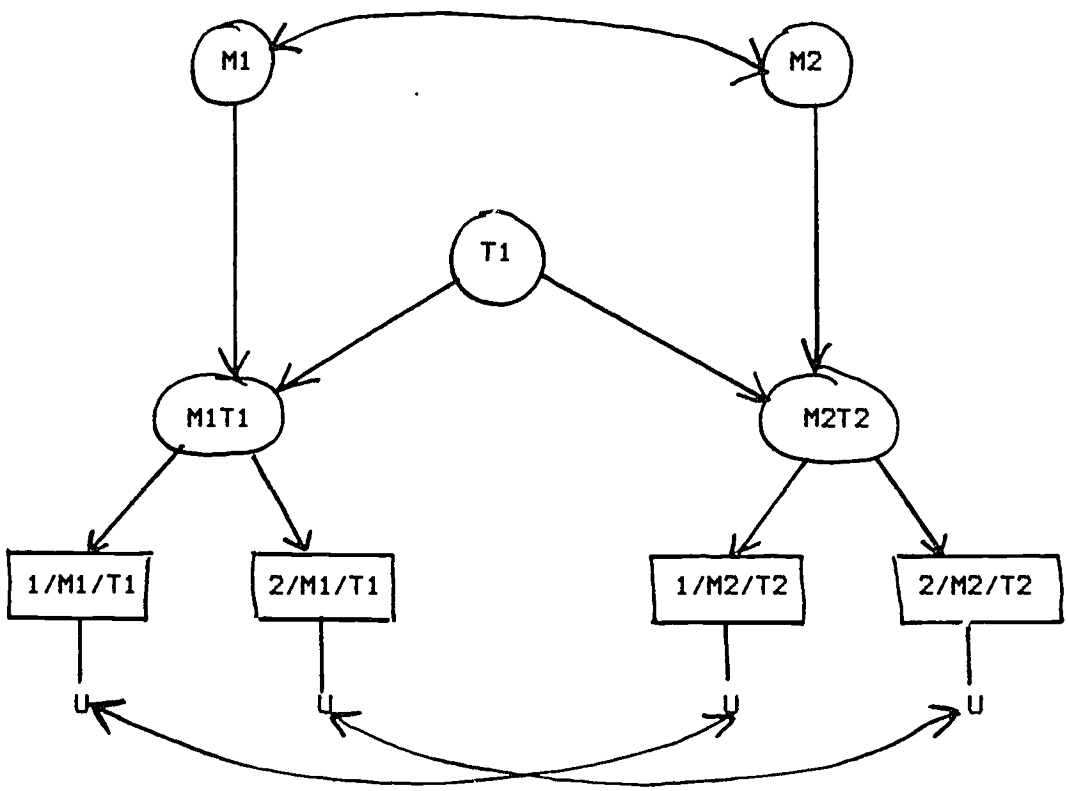

Table 1

First-order Factor Loadings (and Standard Errors) of Each Measured Variable for the General HCFA Model (Model 2)

| Measured | First-order Factors | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Variables | M1T1 | M1T2 | M1T3 | M2T1 | M2T2 | M2T3 | M3T1 | M3T2 | M3T3 |
| 1/M1/T1 | .96/.03 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2/M1/T1 | 1.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3/M1/T1 | .95/.03 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4/M1/T2 | 0 | .98/.02 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5/M1/T2 | 0 | 1.0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6/M1/T2 | 0 | .96/.02 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7/M1/T3 | 0 | 0 | 1.0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8/M1/T3 | 0 | 0 | .84/.[illegible] | 0 | 0 | 0 | 0 | 0 | 0 |
| 9/M1/T3 | 0 | 0 | .99 [illegible]04 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1/M2/T1 | 0 | 0 | 0 | .98/.03 | 0 | 0 | 0 | 0 | 0 |
| 2/M2/T1 | 0 | 0 | 0 | 1.0 | 0 | 0 | 0 | 0 | 0 |
| 3/M2/T1 | 0 | 0 | 0 | .97/.03 | 0 | 0 | 0 | 0 | 0 |
| 4/M2/T2 | 0 | 0 | 0 | 0 | .95/.02 | 0 | 0 | 0 | 0 |
| 5/M2/T2 | 0 | 0 | 0 | 0 | 1.0 | 0 | 0 | 0 | 0 |
| 6/M2/T2 | 0 | 0 | 0 | 0 | .95/.02 | 0 | 0 | 0 | 0 |
| 7/M2/T3 | 0 | 0 | 0 | 0 | 0 | .99/.05 | 0 | 0 | 0 |
| 8/M2/T3 | 0 | 0 | 0 | 0 | 0 | .80/.05 | 0 | 0 | 0 |
| 9/M2/T3 | 0 | 0 | 0 | 0 | 0 | 1.0 | 0 | 0 | 0 |
| 1/M3/T1 | 0 | 0 | 0 | 0 | 0 | 0 | .96/.03 | 0 | 0 |
| 2/M3/T1 | 0 | 0 | 0 | 0 | 0 | 0 | 1.0 | 0 | 0 |
| 3/M3/T1 | 0 | 0 | 0 | 0 | 0 | 0 | .94/.03 | 0 | 0 |
| 4/M3/T2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .98/.02 | 0 |
| 5/M3/T2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.0 | 0 |
| 6/M3/T2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .96/.02 | 0 |
| 7/M3/T3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.0 |
| 8/M3/T3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .82/.05 |
| 9/M3/T3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .99/.04 |

_Note._ See Figure 1 for a description of the measured variables (items) and factors. One variable for each factor was designated to be a reference variable and its unstandardized factor loading was fixed at 1.0. Estimated factor loadings for all other items are presented as a ratio of their standard errors.

Table 2

Correlations Among Latent First-Order Factors From the Target Model (Model 1)

| First-Order Factors | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | M1T1 | 1.00 | | | | | | | | |
| 2 | M1T2 | .58 | 1.00 | | | | | | | |
| 3 | M1T2 | .39 | .01 | 1.00 | | | | | | |
| 4 | M2T1 | .71 | .38 | .31 | 1.00 | | | | | |
| 5 | M2T2 | .38 | .71 | -.07 | .53 | 1.00 | | | | |
| 6 | M2T3 | .31 | .06 | .86 | .39 | .04 | 1.00 | | | |
| 7 | M3T1 | .74 | .46 | .27 | .74 | .44 | .28 | 1.00 | | |
| 8 | M3T2 | .38 | .75 | -.10 | .32 | .69 | .06 | .58 | 1.00 | |
| 9 | M3T3 | .29 | .00 | .87 | .28 | -.02 | .88 | .28 | -.10 | 1.00 |

*Note.* See Figure 1 for a definition of the first-order factors. Correlations are the standardized factor covariances from Model 1 after correction for unreliability due to a lack of internal consistency. Underlined correlations are convergent validities, the correlation between the same trait assessed by two different methods.

Table 3

Parameter Estimates For the General HCFA Model (Model 2) and Standard Errors: Second-Order Factor Loadings (in BETA) and Second-order Factor Variance/Covariances (in PSI)

| | Second-order Factors | | | | | |
|---|---|---|---|---|---|---|
| | M1 | M2 | M3 | T1 | T2 | T3 |
| Factor Loadings (BETA) | | | | | | |
| First-Order Factors | | | | | | |
| M1/T1 | 1.0 | 0 | 0 | .83/.10 | 0 | 0 |
| M1/T2 | .40/.13 | 0 | 0 | 0 | 1.0 | 0 |
| M1/T3 | .19/.03 | 0 | 0 | 0 | 0 | .96/.05 |
| M2/T1 | 0 | 1.0 | 0 | .81/.08 | 0 | 0 |
| M2/T2 | 0 | .45/.15 | 0 | 0 | .97/.05 | 0 |
| M2/T3 | 0 | .23/.09 | 0 | 0 | 0 | .94/.05 |
| M3/T1 | 0 | 0 | .03/.65 | 1.0 | 0 | 0 |
| M3/T2 | 0 | 0 | 1.0 | 0 | .99/.05 | 0 |
| M3/T3 | 0 | 0 | -.01/.12 | 0 | 0 | 1.0 |
| Factor Variance/Covariances (PSI) | | | | | | |
| Second-order Factors | | | | | | |
| M1 | .54/.18 | | | | | |
| M2 | .09/.06 | .41/.13 | | | | |
| M3 | .03/.05 | -.03/.04 | 3.9/76.9 | | | |
| T1 | 0 | 0 | 0 | .84/.11 | | |
| T2 | 0 | 0 | 0 | .42/.06 | .71/.08 | |
| T3 | 0 | 0 | 0 | .27/.05 | -.04/.06 | .76/.08 |

_Note._ See Figure 1 for a description of the first- and second-order factors. Parameters with values of 1.0 were designated to be reference indicators. All other parameter estimates are presented as a ratio their standard errors.

Table 4

The Goodness of Fit For Alternative Models

| Model | $X^2$ (df) | $X^2$/df | RMSR | BBI |
|---|---|---|---|---|
| Uncorrelated Uniquenesses | | | | |
| 0 Null | 12,748 (351) | 36.31 | .449 | .000 |
| 1 Target | 1,563 (288) | 5.43 | .068 | .877 |
| 2 Full | 1,585 (300) | 5.28 | .071 | .876 |
| 3 Trait Only | 1,888 (312) | 6.05 | .083 | .852 |
| 4 Method Only | 2,802 (312) | 8.98 | .281 | .780 |
| Correlated Uniquenesses | | | | |
| 1a Target | 713 (261) | 2.73 | .066 | .944 |
| 2a Full | 736 (273) | 2.70 | .070 | .942 |
| 3a Trait Only | 1,055 (285) | 3.70 | .082 | .917 |
| 4a Method Only | 1,848 (285) | 6.48 | .257 | .855 |

Note. RMSR = Root Mean Square Residual; BBI = Bentler Bonett Index. The purpose of the null model is for computation of the BBI. Each of the other models was first fit with all error/uniquenesses posited to be uncorrelated and then again with correlated error/uniquenesses.