Additive and Multiplicative Method Effects in Applied Psychological Research: An Empirical Assessment of Three Models

Article *in* Journal of Management · June 1994

DOI: 10.1016/0149-2063(94)90006-X

Thomas E. Becker
University of South Florida

Journal of Management
1994, Vol. 20, No. 3, 625-641

Additive and Multiplicative Method Effects in Applied Psychological Research: An Empirical Assessment of Three Models

Thomas E. Becker
Washington State University, Pullman

Joseph A. Cote
Washington State University-Vancouver

Method effects can be additive (independent of trait correlations) or multiplicative (associated with trait correlations). This study is the first to empirically assess the relationship between the nature of method effects and the goodness-of-fit of different latent factor models. Specifically, we examined method effects in 17 published multitrait-multimethod data sets and evaluated the usefulness of confirmatory factor analysis, the direct product approach, and Marsh's correlated uniqueness technique for modeling these effects. While each of the models fit some of the data sets well, Marsh's technique appeared to be generally more effective. Also, Campbell and O'Connell's slope index indicated that additive models (confirmatory factor analysis and the correlated uniqueness approach) were not more likely than a multiplicative model (*the direct product model*) *to provide a good fit to data with additive method effects; nor did a multiplicative model provide a better fit than additive models when method effects were multiplicative.*

Method effects exist when shared methods of measurement are a source of covariance among traits; that is, when the true relationships among variables of interest ("traits") are obscured by the fact that the variables were measured by the same method. Researchers in management, like scientists in other areas, have been concerned with the issue of shared method because knowledge about method effects informs us about the validity of our research instruments. Examples of method effects of concern to researchers in our field include socially desirable responding (Arnold, Feldman & Purbhoo, 1985; Rosenkrantz, Luthans & Hennessey, 1983), halo effects in performance appraisal, and response sets in self-report questionnaires (Spector, 1987).

In order to investigate method effects (as well as to examine other issues of validity), multitrait-multimethod data are invaluable (Campbell & Fiske,

Direct all correspondence to: Thomas E. Becker, Washington State University, Department of Management and Systems, Pullman, WA 99164-4726.

1959). Confirmatory factor analysis (CFA) has emerged as the single most widely recommended and used means for analyzing such data (Joreskog, 1974; Marsh & Hocevar, 1988; Schmitt & Stults, 1986; Widaman, 1985). Typically, researchers assume a single method factor for each measurement method and a single trait factor for each measured construct. Then, using a program such as LISREL (Joreskog & Sorbom, 1989) or EQS (Bentler, 1989), multitrait-multimethod data are used to estimate trait loadings, method loadings, inter-trait correlations, and inter-method correlations. Nested models are often then compared in order to determine the validity of the constructs and the extent of method variance (Widaman, 1985).

Despite the widespread use of CFA for analyzing multitrait-multimethod data, there is currently considerable controversy surrounding use of this technique. For example, recent authors have noted the frequency of problems such as nonconvergence, under-identification, Heywood cases, and other ill-defined solutions resulting from the use of CFA (e.g., Brannick & Spector, 1990; Dillon, Kumar & Mulani, 1987; Marsh, 1989).

In addition, there is a growing concern that the typical use of CFA ignores possible multiplicative trait-method interactions (Browne, 1984; Cudeck, 1988; Schmitt & Stults, 1986; Wothke & Browne, 1990). Multiplicative trait-method interactions exist when the size of the method effects depends on the size of the correlation among the measured traits; for instance, such interactions occur when larger method effects are associated with larger intertrait correlations (Campbell & O'Connell, 1967). The conventional interpretation of confirmatory factor analysis is that this technique assumes that method effects are additive; that is, that the magnitude of method effects is independent of the size of intertrait correlations. Several authors have asserted that because instances of multiplicative trait-method interactions may not be uncommon, the assumption of additive method effects underlying CFA may be called into question (Bagozzi & Yi, 1990, 1991; Bagozzi, Yi & Phillips, 1991; Becker & Vance, 1993).

In response to the problems associated with the use of traditional CFA, a number of other sophisticated models have been developed. Two of these models of particular interest in this study are the direct product (DP) model and the correlated uniqueness (CU) approach. The DP model was developed by Swain (1975) and Browne (1984). This model assumes multiplicative trait-method interactions among measured variables. That is, the model assumes that correlations among variables are associated with method effects. The refined model developed by Browne expresses the covariance (or correlation) matrix of measured variables as the direct product of a covariance (correlation) matrix of traits and a covariance (correlation) matrix of methods, and allows for measurement errors and different scales of measurement among the observed variables.

The CU model was popularized by Marsh (1989). Like CFA, the CU model appears to assume additive method effects. However, rather than representing method effects as explicit factors (as in the case with confirmatory factor analysis), the CU model specifies method effects as correlated uniqueness (i.e., error terms).

Several studies have attempted to compare the above models. Marsh (1989) found that in comparison to CFA, the CU approach seemed to be less prone to ill-defined solutions and the possible confounding of trait and method effects. Marsh and Bailey (1991) found that the CU method converged to proper solutions more often than did CFA and that parameter estimates for this model were more accurate in relation to known population parameters. Kenny and Kashy (1992), in a reanalysis of a previous data set, also demonstrated limitations to CFA and suggested that the CU model may often be a preferable alternative.

Bagozzi, Yi and their colleagues have compared CFA and the DP model. Bagozzi and Yi (1990) reported that for eleven data sets in the literature on affect and perceptions at work, the DP model fit one and CFA fit the other 10. They asserted that if CFA fits the data method effects are necessarily additive, while if the DP model fits method effects must be multiplicative. Further, Bagozzi and Yi concluded, "The results across studies tend to support the premise that MTMM data can be explained by either additive or multiplicative method effects, but not by both" (p. 555). Bagozzi and Yi (1991) examined four multitrait-multimethod data sets in the literature on consumer behavior. They found that the DP model fit two of the four data sets while CFA fit none. The authors concluded that methods often have multiplicative effects and that when multiplicative effects exist, "... the confirmatory factor analysis model will be inappropriate" (p. 429). Finally, Bagozzi, Yi and Phillips (1991) looked at four different data sets from organizational research and found that the DP model fit one data set and CFA fit two. They concluded that the DP model is a solution to the problem of the inability of CFA to test for interactions between traits and methods. They also suggested that if the DP model fits the data then method effects must be multiplicative, and if CFA fits the data then method effects must be additive. Kumar and Dillon (1992), in a conceptual discussion, examined CFA, the DP model, and another technique (covariance components analysis). Their conclusions were quite different from those identified by Bagozzi and Yi (1990, 1991) and Bagozzi, Yi and Phillips (1991). First, Kumar and Dillon asserted that there is no necessary connection between the existence of additive (or multiplicative) effects and appropriateness of additive (or multiplicative) factor models. This is in direct contrast with Bagozzi and Yi's (1990, 1991) assertion that additive models must fit better when method effects are additive while multiplicative models must fit better when method effects are multiplicative. Second, Kumar and Dillon stated that, "Whenever a direct product model is found to fit the data, it is typically possible to find a confirmatory factor analysis model that also provides an adequate fit" (1992, p. 63). This is in opposition to the conclusions of Bagozzi and Yi (1990).

There are a number of limitations to the comparative studies discussed above. First, comparisons of the CU model with CFA have been based largely on simulated data (Marsh, 1989; Marsh & Bailey, 1991, Study 2). While there have been several studies involving "real" data (Bagozzi, 1993; Kenny & Kashy, 1992; Marsh & Bailey, 1991, Study 1), the samples upon which these studies are based have been limited in their representativeness; e.g., Marsh and Bailey (1991, Study 1) used a sample of tenth grade students, and Bagozzi (1993), in a reanalysis of a prior data set, used a sample of undergraduate psychology students. Data from additional samples reflecting other populations are needed before firm conclusions about the relative effectiveness of the CU model and CFA can be drawn.

Second, as described above, empirical comparisons of the DP model and CFA have assumed a priori that there is a relationship between the nature of method effects (additive versus multiplicative) and the fit of the two models. As Kumar and Dillon (1992) pointed out, this assumption is questionable. Thus, what is needed is a comparison of the models which includes a relatively independent assessment of the nature of method effects and the extent of fit of the models.

Third, to this point there is little evidence regarding the relative effectiveness of the DP and CU models. Only Bagozzi (1993) has examined this issue empirically, using the reanalysis of the sample of undergraduates mentioned above. Based upon his findings Bagozzi concluded, "It thus appears that the DP model and the CFA and CU models can fit the same data, although the likelihood of this happening in practice is unknown" (p. 77). Similarly, Kenny and Kashy noted that multiplicative models, "seem promising but there is not as of yet sufficient application of these models for a thorough evaluation of them" (1992, p. 171).

The purpose of this paper is to overcome the limitations just discussed. In accomplishing this goal, we will examine the relative capacity of CFA, the DP model, and the CU model to fit multitrait-multimethod data in the published literature. The contributions of this study to prior research on the effectiveness of the different models are: (1) a large group of data sets will be examined, all of which are based on real (rather than simulated) data, and many of which have not been included in prior comparisons of the three models; and (b) the nature of method effects will be assessed *independent* of model fit. Before describing our study, a brief discussion of the models is in order.

Description of Models

Confirmatory Factor Analysis Model

The general mathematical model for CFA is:

$$\Sigma = \Lambda \, \Phi \, \Lambda' + \Psi^2 \tag{1}$$

where Σ is a population correlation matrix, Λ is a matrix of factor loadings, Φ is a matrix of factor correlations, and Ψ^2 is a diagonal matrix containing uniqueness variances. In practice, an observed multitrait-multimethod correlation matrix is substituted for Σ, and the other matrices are estimated via computer programs such as LISREL (Joreskog & Sorbom, 1989) or EQS (Bentler, 1989).

As Browne (1984) and Widaman (1985) have shown, when the general model is applied to multitrait-multimethod data the model can be decomposed into trait and method components, as follows:

$$\Sigma = \Lambda_T \Phi_T \Lambda_T' + \Lambda_M \Phi_M \Lambda_M' + \Psi^2 \quad (2)$$

where factor loadings for the traits and methods are contained in Λ_T and Λ_M respectively. This statement of the model makes clear why CFA is referred to as an additive model: The observed correlations among measured variables are mathematically defined as a function of the sum of latent traits and latent methods. According to this model, variance in an observed measure is caused by variation due to: (1) differences in trait scores among respondents; (2) differences in systematic biases associated with the given method; and (3) scale-specific biases and random error.

The output matrices of interest for CFA are: (a) lambda, a matrix of factor loadings of the measured variables on the latent variables (i.e., traits and methods); and (b) phi, a matrix of inter-trait correlations. Readers interested in further description of CFA should see Joreskog and Sorbom (1989).

The Direct Product Model

The mathematical model underlying the DP approach is:

$$\Sigma = D_\zeta(P_M \otimes P_T + D_N^2)D_\zeta \quad (3)$$

where D_ζ is a diagonal matrix of communalities, D_N^2 is a diagonal matrix of elements representing nuisance parameters, P_M is a method by method non-negative definite matrix with unity diagonals, P_T is a trait by trait non-negative definite matrix with unit diagonals, and $\otimes$ is the Kronecker (right direct) product. This statement of the model shows why the DP model is referred to as a multiplicative model: The observed correlations among measured variables are mathematically defined as a function of the product between latent traits and latent methods.

The matrices of interest for the DP model are: (a) zeta, a matrix of communalities describing the amount of variance in the measured variables accounted for by the model; (b) rho-m, a matrix of correlations among methods; and (c) rho-t, a matrix of correlations among traits. For further details of this model, see Bagozzi and Yi (1990), Browne (1984), and Wothke and Browne (1990).

The Correlated Uniqueness Model

The general mathematical model underlying Marsh's model is identical to that underlying CFA, to wit:

$$\Sigma = \Lambda \Phi \Lambda' + \Psi^2 \quad (4)$$

However, for the CU model, Λ is a matrix of trait factor loadings (there are no method factors), Φ is a matrix of intercorrelations among traits, and Ψ^2 contains uniqueness variances on the diagonal and covariances among uniqueness off the diagonal. Method effects are inferred on the basis of correlated errors. For further description of this model, see Marsh (1989) and Marsh and Bailey (1991).

METHODS

Cote and Buckley (1987) conducted an exhaustive search (1959-1985) of the multitrait-multimethod literature and identified 70 matrices that they included in their investigation. We examined these same 70 data sets for possible inclusion in our study. In addition, to ensure that all relevant research was included, we conducted a computer search using "multitrait multimethod" and "MTMM" as key words. We then examined this list for the major journals that publish management research. The period of this search was 1980-1991. The journals searched were *Journal of Applied Psychology*, *Academy of Management Journal*, *Personnel Psychology*, and *Organizational Behavior and Human Decision Processes*. This search unearthed an additional 30 data sets.

These 100 data sets were screened on several criteria. First, the published article had to contain the correlation or covariance matrix of measured variables and the matrix had to be positive definite. Second, as recommended by Brannick and Spector (1990), only multitrait-multimethod matrices with more than two different traits and methods were included. Third, matrices with low values on the validity diagonal (average < 0.4) were deleted because convergence was not demonstrated. Finally, as suggested by Boomsma (1982), only data sets with sample sizes greater than (or very close to) 100 were selected. Based on these four criteria, 17 data sets were selected to be included in the present study. The appendix lists the studies from which these data sets were drawn.

EQS version 3.0 (Bentler, 1989) was used to estimate the CFA and CU models. MUTMUM (Browne, 1990) was used to estimate the DP model. For each of the models and data sets, we assumed that one factor existed for each trait and method. In all cases, we allowed correlations among traits and among methods. Also, to provide one basis of comparison for the method-effects models, we ran a CFA model with multiple traits but no methods (i.e., assuming no method effects). Automatic start values were used to fit each model, and if the model failed to converge after 500 iterations the start values were set near estimates from other specifications that converged, and the analysis repeated. We used the following criteria to evaluate the appropriateness of the models:

1. *The number of parameter estimates held at boundary values.* Our decision rule was that, to be appropriate, a model must not have resulted in any boundary estimates. This rule, while stringent, is consistent with the admissibility requirements advocated by statisticians (Wothke, 1987);

2. *Goodness-of-fit of the model to the data.* Because of the problems related to interpreting chi-square values (Bentler & Bonett, 1980), we did not use these values as a criteria for evaluating appropriateness. Instead, to compare the goodness-of-fit across the three models, we calculated the comparative fit index, normed non-centrality and relative non-centrality indices, and the Tucker-Lewis index. Evidence suggests that these measures of fit are unbiased and relatively independent of sample size (Bentler, 1990; Goffin, 1993; McDonald & Marsh, 1990). Our decision rule was that, to indicate a good fit, all four indices had to be greater than .90.

Finally, for each of the 17 data sets included in this study, we calculated the index of the nature of method effects developed by Campbell and O'Connell (1967). This index, called the slope, is the ratio of the variance of the heterotrait-monomethod correlations to the variance of the heterotrait-heteromethod correlations. If method effects are additive, the monomethod correlations will be equal to the heteromethod correlations plus some constant bias (MM=HT+c). In addition, the variance for these two terms would be equal [var(MM) = var(HT+c)]. Because c is a constant, it can be eliminated from the variance equation [var(HT+c) = var(HT)]. Therefore, if method effects are additive the slope (var(HT)/var(MM)) will be 1.0, (i.e., the variance of the heterotrait-monomethod correlations will equal the variance of the heterotrait-heteromethod correlations). However, if method effects are multiplicative, the slope will be greater than 1.0 (i.e., the variance of the heterotrait-monomethod correlations will be greater than the variance of the heterotrait-heteromethod correlations).

One potential limitation to Campbell and O'Connell's index is that it does not take into account error of measurement inherent in the observed correlations. To address this issue, we decided to correct (wherever possible) the observed correlations for unreliability prior to computing the slope index.

Results

Table 1 reports the sample size, number of traits and methods, and slope for each of the data sets. This table also reports the results of the null model for each data set. To determine whether method effects were additive or multiplicative, we first used estimates of reliability provided by the original researchers to correct the observed correlations for attenuation; as shown in Table 1, this was possible for 10 of the 17 data sets. Next, we computed the point-estimates for the slope for each set of correlations. Finally, we set up a 95% confidence interval around each slope. We designated those data sets whose slope interval contained 1.0 as additive and those whose did not as multiplicative. Based on this analysis, 12 of the data sets demonstrated additive effects and five demonstrated multiplicative effects.

Table 1. Description of Data Sets and Results for the Null Model

Study	*n*	*Traits*	*Methods*	*Null Model* *df*	*Null Model* χ^2	*Mean Slope*	*95% CI*	
Allen	177	3	3	36	1083.3	0.900	0.46,	1.34
Arora	96	3	3	36	571.7	0.711	0.42,	1.00[a]
Dunham	622	4	4	120	6799.7	0.897	0.51,	1.29[a]
Elbert	250	4	3	66	2277.9	1.272	1.19,	1.35[a,b]
Flamer #1	105	3	3	36	346.0	1.183	0.98,	1.39[a]
Flamer #2	105	3	3	36	405.7	1.232	0.88,	1.59[a]
Freedman	149	5	3	105	1542.4	0.995	0.85,	1.14
Hicks	119	3	3	36	508.7	2.107	1.34,	2.87[a,b]
Kothandapani	100	3	4	66	785.5	1.272	1.03,	1.52[a,b]
Marsh #1	286	5	3	105	5711.1	1.276	0.88,	1.67
Marsh #2	392	5	3	105	6375.1	1.093	0.70,	1.49
Meier	320	3	3	36	1682.2	1.076	0.99,	1.17[a]
Ostrom	189	3	4	66	1871.9	0.817	0.60,	1.04
Roberts	120	3	3	36	1029.0	1.174	0.72,	1.63[a]
Schmitt	310	7	4	378	11320.8	1.063	1.04,	1.09[b]
Seymour	132	4	3	66	1698.0	0.921	0.88,	0.97
Shavelson	99	3	3	36	482.3	6.069	1.36,	10.78[a,b,c]

Notes: [a] Correlations in MTMM matrix corrected for attenuation.
[b] Slope index indicates that method effects are multiplicative.
[c] The variances for one heteromethod triangle was 0. This triangle was deleted from the analysis.

Table 2. Confirmatory Factor Analysis Results: Traits Only[a]

Study	*df*	χ^2	*CFI*	*NNCI*	*RNCI*	*TLI*	*Boundary Estimates*
Allen	24	205.6	.83	.60	.83	.74	0
Arora	24	107.6	.84	.65	.84	.77	0
Dunham	98	843.8	.74	.55	.89	.86	0
Elbert	48	122.6	.97	.86	.97	.95	0
Flamer #1	24	23.3	.98	1.00	1.00	1.00	0[b]
Flamer #2	24	31.1	.98	.97	.98	.97	1
Freedman	80	575.8	.66	.19	.66	.55	1
Hicks	24	138.1	.76	.62	.76	.64	3
Kothandapani	51	387.3	.55	.19	.53	.40	1
Marsh #1	80	1353.9	.77	.11	.77	.70	6
Marsh #2	80	1250.5	.81	.22	.81	.75	4
Meier	24	759.3	.69	.32	.55	.33	0
Ostrom	51	135.5	.95	.80	.95	.94	0
Roberts	24	188.8	.83	.50	.83	.75	0
Schmitt	329	1018.6	.94	.33	.94	.93	0
Seymour	48	259.4	.87	.45	.87	.82	0
Shavelson	24	122.8	.78	.61	.79	.67	0

Notes: [a] CFI = comparative fit index, NNCI = normed non-centrality index, RNCI = relative non-centrality index, and TLI = Tucker-Lewis index
[b] Model meets the criteria of good fit and no boundary estimates.

Table 3. Confirmatory Factor Analysis Results: Traits and Methods

Study	*df*	χ^2	*CFI*	*NNCI*	*RNCI*	*TLI*	*Boundary Estimates*	r^a
Allen	12	12.3	1.00	.99	.99	.99	1	.78
Arora	12	17.0	.99	.97	.99	.97	0^b	.61
Dunham	76	260.0	.97	.86	.97	.96	1	.50
Elbert	33	36.8	.99	.99	.99	.99	1	.39
Flamer #1	12	5.1	1.00	1.03	1.02	1.07	2	.83
Flamer #2	12	11.0	1.00	1.00	1.00	1.01	3	.72
Freedman	62	158.9	.93	.72	.93	.89	1	.66
Hicks	12	48.2	.92	.86	.92	.77	3	.37
Kothandapani	33	52.7	.97	.91	.97	.95	1	.27
Marsh #1	62	224.3	.97	.75	.97	.95	1	.52
Marsh #2	62	220.5	.98	.82	.97	.96	2	.83
Meier	12	11.8	1.00	1.00	1.00	1.00	2	.54
Ostrom	33	22.4	1.00	1.03	1.01	1.01	1	.85
Roberts	12	46.3	.97	.87	.97	.90	3	.61
Schmitt	295	465.2	.98	.76	.98	.98	0	.92
Seymour	33	89.1	.97	.81	.97	.93	1	.34
Shavelson	12	5.6	1.00	1.03	1.01	1.04	1	.49

Notes: [a] Average estimated correlation among methods (based on absolute values).
[b] Model meets the criteria of good fit and no boundary estimates.

Table 2 contains the degrees of freedom, chi-squares, measures of fit, and number of boundary estimates for the CFA traits-only model. This model satisfies our criteria for one of the data sets (Flamer #1); thus, it appears that there are little if any effects of method for this data. The traits-only model also fits a number of other data sets without producing boundary estimates. However, the goodness-of-fit measures indicate a relatively poor fit of the traits-only model to these data.

Table 3 contains the results for the CFA model with both traits and methods. As can be seen, using our criteria this model is appropriate for only one of the 17 data sets (Arora). The goodness-of-fit indices for this data set suggest that the model with multiple methods fits considerably better than the traits-only model. The slope index suggests that method effects for the Arora data set are additive. Table 3 also reports the average method correlation for the data sets. This information will be useful in evaluating the results of the CU model.

Table 4 reports the results for the DP model. Our criteria suggest that this model is appropriate for four of the data sets, none of which were adequately fit by CFA with traits and methods. However, it might be noted that the traits-only model, while not meeting our goodness-of-fit criteria, does seem to provide a fair fit to two of these data sets (Elbert and Ostrom). The slope index indicates that two of the four data sets for which the DP model is appropriate contain additive method effects while the other two contain multiplicative effects.

Table 5 shows the results for the CU model. This model fares better than the others, being appropriate for seven of the 17 data sets: the one for which

Table 4. Direct Product Model Results

Study	*df*	χ^2	*CFI*	*NNCI*	*RNCI*	*TLI*	*Boundary Estimates*
Allen	21	62.9	.96	.89	.96	.93	3
Arora	21	46.9	.95	.87	.95	.92	6
Dunham	92	399.3	.95	.78	.95	.94	2
Elbert	45	49.8	.99	.99	.99	.99	0[a]
Flamer #1	21	22.4	.99	.99	.99	.99	1
Flamer #2	21	25.6	.99	.98	.99	.98	3
Freedman	77	244.2	.86	.57	.88	.84	1
Hicks	21	85.8	.86	.76	.86	.77	6
Kothandapani	45	122.1	.96	.68	.89	.84	5
Marsh #1	77	324.4	.96	.65	.96	.94	1
Marsh #2	77	413.0	.95	.65	.95	.93	1
Meier	21	45.5	.99	.96	.99	.97	0[a]
Ostrom	45	58.1	.99	.97	.99	.99	0[a]
Roberts	21	166.1	.85	.55	.85	.75	2
Schmitt		Data exceeds MUTMUM program size limits					
Seymour	45	259.2	.87	.44	.87	.81	2
Shavelson	21	22.3	.99	.99	.99	.99	0[a]

Note: [a] Model meets the criteria of good fit and no boundary estimates. Note that the Elbert and Shavelson data sets have a multiplicative slope index while the Meier and Ostrom data sets have additive indices.

Table 5. Correlated Uniqueness Model Results

Study	*df*	χ^2	*CFI*	*NNCI*	*RNCI*	*TLI*	*Boundary Estimates*
Allen	15	26.2	.99	.97	.99	.97	0[a]
Arora	15	21.7	.99	.97	.99	.97	0[a]
Dunham	61	753.6	.90	.57	.90	.80	0
Elbert	30	24.0	1.00	1.01	1.00	1.01	0[a]
Flamer #1	15	14.9	1.00	1.00	1.00	1.00	0[a]
Flamer #2	15	12.2	1.00	1.01	1.01	1.02	1
Freedman	50	180.1	.91	.65	.91	.81	3
Hicks	15	25.7	.98	.96	.98	.95	2
Kothandapani	39	71.6	.96	.85	.95	.92	0
Marsh #1	50	169.7	.98	.81	.98	.96	1
Marsh #2	50	249.5	.97	.78	.97	.93	2
Meier	15	30.7	.99	.98	.99	.98	0[a]
Ostrom	39	53.8	.99	.96	.99	.99	0[a]
Roberts	15	113.6	.90	.66	.90	.76	0
Schmitt	245	391.9	.95	.79	.99	.98	0
Seymour	30	110.6	.95	.74	.95	.89	0
Shavelson	15	12.5	1.00	1.01	1.01	1.01	0[a]

Note: [a] Model meets the criteria of good fit and no boundary estimates. Note that the Allen, Arora, Flamer, Meier, and Ostrom data sets have an additive slope index, while the Elbert and Shavelson data sets have a multiplicative index.

the traits-only CFA was also appropriate, the one for which the traits and methods CFA was also appropriate, the four for which the DP model was also appropriate, and one other (Allen). The goodness-of-fit measures for the Allen data set are substantially higher that the corresponding indices in the traits-only model. All in all, at least one of the three models was appropriate for seven of the data sets, while none of the models were appropriate for the other ten.

To compare the goodness-of-fit of the models we examined the comparative fit indices. Descriptively, the CFA had the best fit for 10 of the 17 data sets, while the CU model fit best in 4 cases. Both the CFA and CU models fit equally well in 3 cases. The DP model did not fit better than the CFA or CU models for any of the studies. A Kruskal-Wallis one-way ANOVA indicated that both the CFA and CU models fit better than did the DP model, $p < .05$. The fit of the CFA model and CU model were not statistically different.

Discussion

Interpretation of Findings

With respect to the usefulness of the models, our results support three conclusions. First, based upon our criterion for admissibility (i.e., models with boundary values were ruled inappropriate), the CU model seems to have fewer estimation problems than either CFA (with traits-and-methods) or DP model. This supports the finding of Marsh (1989) and Marsh and Bailey (1991), and goes beyond the prior research by showing that the CU model holds up well in comparison to the direct product model and for a large variety of samples. This is not to imply that the CU model is faultless. As others have noted (Bagozzi, 1993; Kenny & Kashy, 1992), a potential problem with the CU model is that it assumes methods are uncorrelated. The estimates of method correlations supplied by CFA (in Table 2) suggest that this assumption was violated for the data sets we examined. This indicates that the CU model supplies biased estimates of convergent and discriminant validity for these data sets.[1]

Our second conclusion is that, estimation problems aside, CFA (with traits and methods) tends to fit multitrait-multimethod matrices somewhat better than do the DP and CU models. This leads us to recommend that if boundary estimates are not a problem, CFA is generally preferable to alternative methods. Nevertheless, the goodness-of-fit indices for data included in this study suggest that where the CFA provides a good fit, the DP and CU models also appear to fit well.

Third, our results suggest that the DP model is generally the weakest of the three approaches. Whenever the DP approach resulted in an appropriate solution, the CU model also produced a satisfactory solution. In addition, both the CFA traits-and-methods and CU models tended to provide better fits to the data. Still, we cannot rule out the possibility that the DP model may be more appropriate than additive models under some circumstances. Further, it should be noted that consistent with prior research (Bagozzi, 1990, 1991), when the DP model fit the data, CFA did not.

Most researchers who have argued for the usefulness of multiplicative models have cited the 1967 article by Campbell and O'Connell (Bagozzi & Yi, 1990; 1991; Bagozzi, Yi & Phillips, 1991; Browne, 1984; Cudeck, 1988; Swain, 1975; Wothke & Browne, 1990), who identified additive and multiplicative effects using the slope index. Later authors then assumed that when multiplicative method effects exist, the direct product model is more appropriate than an additive model. To understand why this conclusion may be false, the words of Kumar and Dillon are enlightening:

> Campbell and O'Connell (1967, 1982) use the term "factor" in an ANOVA (analysis of variance) sense; that is, the method facet is viewed as an independent categorical variable introduced to account for variations in the values of the observed *correlations*. Elsewhere, however, they use the term "factor" to denote, in the classic factor analytic sense, latent variables that are introduced explicitly to account for the covariation, as opposed to variation, among a set of observed *measures*. Their apparent dual use of the term "factor" obfuscates the fact that there is no necessary connection between (1) the additive (or multiplicative) effects observed at the level of correlations (or covariances) between trait-method pairings and (2) the additive or multiplicative factor models for observed scores on those pairings (1992, p. 53).

Our results indicate that the appropriateness of additive and multiplicative models appear to be unrelated to Campbell and O'Connell's slope index. This supports Kumar and Dillon's (1992) argument that there is no necessary connection between the nature of method effects (additive versus multiplicative) and the appropriateness of additive or multiplicative models. In addition, while the slope index and findings regarding model appropriateness often suggested different conclusions about the nature of method effects, both sets of findings indicated that additive effects are much more common than multiplicative effects. This is contrary to Campbell and O'Connell's (1982) contention that multiplicative method effects are pervasive.

Kumar and Dillon (1992) also conjectured that whenever the DP model is appropriate and fits a given data set, a CFA solution can also be obtained. However, we found that for the four cases in this study where the DP model was appropriate the analogous CFA solution (i.e., assuming one factor for each trait and method) was not. Thus, while it is possible that some other model (e.g., with a different number of traits or methods) may have proven to satisfy our criteria of appropriateness, it does not seem to be true that an analogous CFA solution can be found whenever a DP model is appropriate. It should be remembered, though, that an appropriate and analogous CU model solution *was* obtained for all the cases for which the DP model was appropriate. These findings are contrary to the suggestions of Bagozzi and Yi (1990, 1991) and Bagozzi, Yi, and Phillips (1991), who argued that additive and multiplicative models should not fit the same data.[2] Therefore, perhaps it is often correct that when a multiplicative model is appropriate some appropriate additive model can also be found. Future research should investigate this issue further.

One disturbing finding is that while each of the models fit some of the data sets well, none of the models fit 10 of the 17 matrices. Seven of these ten data sets evinced additive slopes and three demonstrated multiplicative slopes. Therefore, contrary to the assertions of prior authors (Bagozzi & Yi 1990, 1991; Bagozzi, Yi & Phillips 1991; Browne, 1984; Wothke & Browne, 1990), it does not appear that estimation problems of additive models are largely due to the multiplicative nature of method effects. Other authors have suggested that boundary estimates are caused by identification problems (Brannick & Spector, 1990; Kenny & Kashy, 1992). Because the DP and CU models were developed to avoid problems of statistical under-identification (Browne, 1984; Marsh, 1989), it seems unlikely to us that identification problems were the cause of improper solutions in our study. In short, the issue of whether boundary estimates are caused by model misspecification, flaws in the data, or some complex interaction between the two remains something of an enigma.

Implications for Management Research

Based upon this study, we can make several recommendations to those doing research in the area of management. While selecting multiple methods of measuring a construct, we urge management researchers to identify methods that are unlikely to be artifactually correlated. This will increase the chances of arriving at a good fitting CU model without violation of the model's assumption of uncorrelated methods. Observed correlations among methods are due to at least three factors: (1) measurement of a common trait (which promotes convergent validity); (2) measurement of a common artifact (which promotes method variance); and (3) random error. For instance, consider a study where employee performance is measured by ratings from two levels of supervision: the first level supervisor and his or her boss. The correlation between ratings may reflect the shared observation of actual performance (common trait), shared bias to rate higher those employees who are ingratiating to management (common artifact), and random error. Holding common trait variance equal, the smaller the shared bias in ratings the less correlated will be the methods and, hence, the more appropriate will be the CU model. Therefore, selecting two sources of ratings (e.g., supervisors and co-workers) which are less likely to share a given bias would be a wiser choice.[3]

Method effects can also be reduced while implementing a study. For example, in research on the faking of biodata in employee selection, empirical and option keying appears to reduce the influence of socially desirable responding (Haymaker, 1986; Kluger, Reilly & Russell, 1991), as does warning job applicants that their answers will be verified (Hough, Eaton, Dunnette, Kamp & McCloy, 1990). Our results underline the importance of such research tactics. Anything that can be done to reduce artifactual covariance among methods will increase the probability of arriving at a satisfactory analytic solution.

A final implication of our study involves the development of a rational strategy for analyzing data from management studies. Bagozzi (1993) recommended that the analysis of multitrait-multimethod data begin with an examination of the Campbell and Fiske (1959) criteria. Second, regardless of whether or not these criteria are met, Bagozzi suggested that researchers attempt to run the CFA and CU models. Next, if additive models do not fit the data well or results in improper solutions, Bagozzi recommended that researchers apply the DP model. Bagozzi's logic for running the DP model after the additive models was that the additive models are conceptually simpler and, if they fit well, easier to explain.

Given our results, we would amend Bagozzi's (1993) strategy as follows. We agree that researchers should attempt to run CFA prior to the DP model. CFA tends to fit multitrait-multimethod matrices better and is probably somewhat easier to interpret and explain. Widaman's (1985) approach for comparing hierarchically nested models allows researchers to assess the relative fit of traits-only and traits-and-methods models within the context of CFA. However, we suggest that if CFA does not produce appropriate results, the DP model should be run next (i.e., *prior* to the CU model), using the hierarchical approach discussed by Becker and Vance (1993). Because the assumption of uncorrelated methods may often be violated, interpretation of the results from the CU model may be most difficult of all. Thus, we recommend that the final step in the analysis strategy involve running the CU model (if neither the CFA nor DP models work). If the CU model fits, the researcher must consider the issue of correlations among methods in his or her discussion. If the researcher has heeded our above advice regarding the selection of methods and reduction of method variance through study design, this last step may be much less painful.

Limitations and Future Directions

Several limitations to our investigation deserve comment. First, the slope index that we used to assess the additivity and multiplicativity of method effects is only a rough measure of the nature of such effects. It is likely that the nature of method effects is more complicated than is recognized by the dichotomy of additivity versus multiplicativity. Also, while we attempted to improve the usefulness of the slope index by correcting the observed correlations for unreliability, this correction was not possible in all cases. Some of the original articles simply did not provide information on the reliability of measures. Despite these problems we would point out that, for our purposes, even a rough approximation of the nature of method effects was better than none. Also, we note that correcting for attenuation does not appear to substantively change our central conclusions (i.e., that the CU model fits whenever either of the other two models fit, and that the fit of the models is not closely tied to the nature of method effects).

Second, for all three approaches, we assumed that one factor existed for each trait and for each method. Thus, we only examined one model within each generic framework. While this allowed us to compare analogous latent structures for the three models, it may be that our results would have differed had we made different assumptions. Our rationale for assuming one factor for each trait and method was that: (1) all else held constant, this latent structure will fit as well or better than more restrictive models; and (2) as long as the same assumption was used for all three approaches the comparisons would be fair. Future research comparing CFA, CU, and DP models should include a wider variety of specific models (e.g., with different numbers of traits, methods, or both).

Finally, one could argue that our criteria for evaluating the appropriateness of the models were harsh. We agree that many researchers have used more liberal criteria, such as allowing one or more boundary estimates. Nevertheless, in most cases it seems that different criteria would not dramatically change our conclusions. For instance, even if one allows those solutions with one boundary estimate to be defined as appropriate, the interpretation of most of our results does not change: Wherever the DP model fits an additive model also fits, the CU model still seems to have fewer estimation problems than the other models, and so on. At any rate, we applied the same criteria across all three models; this seems to us to be an impartial means of evaluation.

In conclusion, this study contributes to past research by addressing several important issues not addressed in prior empirical comparisons of additive and multiplicative models. Future research should extend this study by examining multitrait-multimethod matrices not contained in this article, and by comparing the models given different sets of assumptions and different criteria of evaluation.

Acknowledgment: We thank Richard Bagozzi, Michael Browne, Ajith Kumar, and this journals reviewers for comments on earlier versions of this article.

Appendix

The data sets included in this study were drawn from the following published articles:

Allen, J. G. & Hamsher, J. H. (1974). The development and validation of a test of emotional styles. *Journal of Consulting and Clinical Psychology, 42*: 663-668.

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

Dunham, R., Smith, F. & Blackburn, R. (1977). Validation of the index of organizational reactions with the JDI, MSQ, and the faces scales. *Academy of Management Journal, 20*: 420-432.

Elbert, N. (1979). Questionnaire validation by confirmatory factor analysis:
An improvement over multitrait-multimethod matrices. *Decision Sciences, 10*: 629-644.

Flamer, S. (1983). Assessment of the MTMM matrix validity of likert scales via confirmatory factor analysis. *Multivariate Behavioral Research, 18*: 275-308.

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.

Hicks, J. M. (1967). Comparative validation of attitude measures by the multitrait-multimethod matrix. *Educational and Psychological Measurement, 27*: 985-995.

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.

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

Meier, S. (1984). The construct validity of burnout. *Journal of Occupational Psychology, 57*: 211-219.

Ostrom, T. (1969). The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Applied Social Psychology, 5*: 12-30.
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.
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.
Seymour, D. & Lessne, G. (1984). Spousal conflict arousal: Scale development.*Journal of Consumer Research, 11*: 810-821.
Shavelson, R. J. & Bolus, R. (1982). Self concept: The interplay of theory and methods. *Journal of Educational Psychology, 74*: 3-17.

Notes

1. One might note that the CFA traits-only model also seems to have fewer estimation problems than the CFA traits-and-methods and DP models. However, the traits-only model typically provided a relatively poor fit to the data sets in this study. This suggests that method effects are quite widespread in applied psychological research.
2. In all fairness, we wish to acknowledge that Bagozzi (1993), in later work, has recognized the possibility that additive and multiplicative models may sometimes fit the same data.
3. To the extent that two methods are highly correlated due to measurement of a common trait (rather than due to measurement of a common artifact), the CFA traits-only model should provide a good fit.

References

Arnold, H. J., Feldman, D. C. & Purbhoo, M. (1985). The role of social desirability response bias in turnover research. *Academy of Management Journal, 28*: 955-966.
Bagozzi, R. P. (1993). Assessing construct validity in personality research: Applications to measures of self-esteem. *Journal of Research in Personality, 27*: 49-87.
Bagozzi, R. P. & Yi, Y. (1990). Assessing method variance in multitrait- multimethod matrices: The case of self-reported affect and perceptions at work. *Journal of Applied Psychology, 75*: 547-560.
________. (1991). Multitrait-multimethod matrices in consumer research. *Journal of Consumer Research, 17*: 426-439.
Bagozzi, R. P., Yi, Y. & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly, 36*: 421-458.
Becker, T. E. & Vance, R. J. (1993). Construct validity of three types of organizational citizenship behavior: An illustration of the direct product model with refinements. *Journal of Management, 19*: 663-682.
Bentler, P. M. 1989. *EQS structural equations program manual.* Los Angeles: BMDP Statistical Software, Inc.
________. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*: 238-46.
Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*: 588-606.
Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analytic models. Pp. 149-173 in K.G. Joreskog & J. Wold (Eds.), *Systems under indirect observation: Causality structure prediction.* New York: Elsevier.
Brannick, M. T. & Spector, P. E. (1990). Estimation problems in the block-diagonal model of the multitrait-multimethod matrix. *Applied Psychological Measurement, 14*: 325-339.
Browne, M. W. (1984). The decomposition of multitrait-multimethod matrices. *British Journal of Mathematical and Statistical Psychology, 37*: 1-21.
________. (1990). MUTMUM PC user's guide. Unpublished manuscript, The Ohio State University.
Campbell, D. T. & Fiske, D. (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.
________. (1982). Methods as diluting trait relationships rather than adding irrelevant systematic variance. Pp. 93-111 in D. Brinberg & L. Kidder (Eds.), *Forms of validity in research: New directions for methodology of social and behavioral science,* Vol.12. San Francisco: Jossey-Bass.
Cote, J. A. & Buckley, M. R. (1987). Estimating trait, method, and error variance: Generalizing across seventy construct validation studies. *Journal of Marketing Research, 26*: 315-319.

Cudeck, R. (1988). Multiplicative models and MTMM matrices. *Journal of Educational Statistics, 13*: 131-147.

Dillon, W. R., Kumar, A. & Mulani, N. (1987). Offending estimates in covariance structure analysis: Comments on the causes of and solutions to Heywood cases. *Psychological Bulletin, 101*: 126-135.

Goffin, R. D. (1993). A comparison of two new indices for the assessment of fit of structural equation models. *Multivariate Behavioral Research, 28*: 205-214.

Haymaker, J. C. (1986). Biodata as a predictor of employee integrity and turnover. Paper presented at the 94th Annual Convention of the American Psychological Association, Washington, DC.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D. & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*: 581-595.

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

Joreskog, K. G. & Sorbom, D. (1989). *LISREL 7 user's reference guide*, 1st ed. Mooresville, IN: Scientific Software.

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

Kluger, A. N., Reilly, R. R. & Russell, C. J. (1991). Faking biodata tests: Are option-keyed instruments more resistant? *Journal of Applied Psychology, 49*: 889-896.

Kumar, A. & Dillon, W. R. (1992). An integrative look at the use of additive and multiplicative covariance structure models in the analysis of MTMM data. *Journal of Marketing Research, 29*: 51-64.

Marsh, H. W. 1989. Factor analysis of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13*: 335-361.

Marsh, H. W. & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement, 15*: 47-70.

Marsh, H. W. & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology, 73*: 107-117.

McDonald, R. P. & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin, 107*: 247-255.

Rosenkrantz, S. A., Luthans, F. & Hennessey, H. W. (1983). Role conflict and ambiguity scales: An evaluation of psychometric properties and the role of social desirability response bias. *Educational and Psychological Measurement, 43*: 957-970.

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

Spector, P. E. (1987). Method variance as an artifact in self-reported affect and perceptions at work: Myth or significant problem? *Journal of Applied Psychology, 72*: 438-443.

Swain, A. J. (1975). Analysis of parametric structures for variance matrices. Unpublished doctoral dissertation, University of Adelaide, Australia.

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

Wothke, W. (1987). Multivariate linear models of the multitrait-multimethod matrix. Paper presented at the meeting of the American Educational Research Association, Washington, DC.

Wothke, W. & Browne, M. W. (1990). The direct product model for the MTMM matrix parameterized as a second order factor analysis model. *Psychometrika, 55*: 255-262.