Multivariate Behavioral Research

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hmbr20

Confirmatory Measurement Model Comparisons Using Latent Means

Roger E. Millsap & Howard Everson

Published online: 10 Jun 2010.

Multivariate Behavioral Research, *26* (3), 479-497

Confirmatory Measurement Model Comparisons Using Latent Means

Roger E. Millsap
Baruch College, City University of New York

Howard Everson
Office of Academic Affairs, City University of New York

Confirmatory factor analysis (CFA) is often used to verify measurement models derived from classical test theory: the parallel, tau-equivalent, and congeneric test models. In this application, CFA is traditionally applied to the observed covariance or correlation matrix, ignoring the observed mean structure. But CFA is easily extended to allow nonzero observed and latent means. The use of CFA with nonzero latent means in testing six measurement models derived from classical test theory is discussed. Three of these models have not been addressed previously in the context of CFA. The implications of the six models for observed mean and covariance structures are fully described. Three examples of the use of CFA in testing these models are presented. Some advantages and limitations in using CFA with nonzero latent means to verify classical measurement models are discussed.

Introduction

One important application of confirmatory factor analysis (CFA) lies in the verification of measurement models derived from classical test theory (Alwin & Jackson, 1980; Dwyer, 1983; Hattie, 1985; Jöreskog, 1971, 1978; Kenny, 1979; Loehlin, 1987; McDonald, 1985). Three measurement models are commonly discussed in the literature: the parallel, tau-equivalent, and congeneric models. Each model entails specific restrictions on the common factor model. CFA allows the investigator to impose these restrictions and test the fit of the resulting model. Traditionally, CFA is applied to the observed covariance matrix in fitting the measurement model. As a result, mean differences among the observed variables are ignored and do not influence the fit. But the classical measurement assumptions have implications for mean structures as well as covariance structures, and hence the observed mean structure is relevant. CFA can be formulated to incorporate

---

We extend thanks to John Nesselroade for the use of the data presented in one of the examples, and to David Rindskopf for comments on an earlier draft of this article. Correspondence concerning this article should be addressed to Roger E. Millsap, Department of Psychology, Baruch College, City University of New York, 17 Lexington Ave., New York, NY 10010.

nonzero means on both observed and latent variables (Jöreskog & Sörbom, 1985; Sörbom, 1974, 1978, 1982). Within this formulation, we can impose restrictions on the latent mean structure to reflect measurement model assumptions.

This article describes the use of CFA with nonzero latent means in discriminating among six different measurement models arising from classical test theory. Three of these models have been discussed previously in the context of CFA, but the remaining three models have not been addressed previously. We begin with a brief review of the common factor model with latent means, followed by a description of the six measurement models. The implications of the six models for observed mean and covariance structures are fully described. Three examples of the application of these models in real data are presented, and some further applications are discussed.

Factor Analysis with Latent Means

Let **X** be a $p \times 1$ vector of observable random variables. For convenience, the common factor model is traditionally described under the assumption of null or zero means for the observed and latent variables. Nonzero means are easily incorporated however. The common factor model with nonzero latent means assumes that there exist $m < p$ common factor or latent variables, arrayed in an $m \times 1$ vector ξ, such that

$$\mathbf{X} = \nu + \Lambda\xi + \delta, \tag{1}$$

where ν is a $p \times 1$ vector of intercept parameters, Λ is a $p \times m$ factor pattern matrix, and δ is a $p \times 1$ vector of unique factor variables. Let E() be the mathematical expectation operator. We assume that $E(\delta) = 0$ and $E(\delta\delta') = \theta$, a $p \times p$ diagonal matrix: the unique factors have zero means and are mutually uncorrelated. Additionally, the common factor variables are assumed to be uncorrelated with the unique factor variables. Let these common factor variables have nonzero expectations $E(\xi) = \kappa$, an $m \times 1$ vector, and let Σ be the $p \times p$ covariance matrix for **X**. From previous assumptions,

$$\Sigma = \Lambda\Phi\Lambda' + \theta \tag{2}$$

$$E(\mathbf{X}) = \nu + \Lambda\kappa \tag{3}$$

where Φ is the $m \times m$ common factor covariance matrix. Equation 3 provides the link between the observed mean structure and the common factor model. Given a fixed factor pattern matrix, the parameter vectors ν and κ are not identified in this equation. The addition of any constants to κ can be offset by corresponding subtraction from v. The common factor model with latent means poses identification problems beyond those present in the zero-means model. Specific solutions to these problems are discussed below.

If p observed variables are analyzed, the degrees of freedom for a CFA model with nonzero latent means will be

$$df = (1/2)(p + 2)(p + 1) - t - 1, \tag{4}$$

where t is the number of independent parameters to be estimated. An additional p degrees of freedom are available in the CFA with nonzero latent means because the p observed means are included in the moment matrix to be analyzed. If multiple group data are analyzed, the df in Equation 4 can be calculated separately in each group and summed across groups to give the total df for the model.

Additional assumptions are generally required if CFA is used in testing measurement models from classical test theory. First, the statistical tests of fit associated with maximum likelihood factor analysis require distributional assumptions. A sufficient set of assumptions is that the common and unique factor variables have multivariate normal distributions in the population under study. Secondly, we must clarify the relationship between the unique factor variable and the error score in classical test theory. As noted by several authors (Alwin & Jackson, 1980; Lord & Novick, 1968; McDonald, 1985; Mulaik, 1972), the two sets of scores have similar statistical properties, but need not be identical. Lack of identity can lead to problems in using CFA to verify classical measurement assumptions, and we return to this issue below. Until then, we assume the unique factor and error score variables to be identical. Finally, each of the measurement models to be discussed assumes a single common factor underlying the p observed variables in X. All of the models can be extended to the case of multiple sets of observed variables, with each set based on a single common factor. Models which allow individual observed variables to be determined by more than one common factor (e.g., multitrait-multimethod models) will not be considered.

Classical Measurement Models

Parallel Models

The concept of parallelism among measurements is fundamental in classical test theory (Lord & Novick, 1968). Parallel measures are defined to have identical true scores and identical error score variances. In CFA with zero latent means, the parallel model is implemented by requiring equal factor loadings for all observed variables and equal unique variances (Jöreskog, 1971). These restrictions yield the covariance structure

$$\Sigma = \phi \mathbf{1}\mathbf{1}' + \theta \mathbf{I} \tag{5}$$

with $\mathbf{1}$ a $p \times 1$ unit vector, $\mathbf{I}$ a $p \times p$ identity matrix, ϕ the (scalar) common factor variance, and θ the (scalar) unique factor variance. If the common factor loading is fixed to one for identification, only these two parameters are estimated.

In a CFA with nonzero latent means, the parallel model can be represented by choosing ν to be a null vector in Equation 3, fixing all factor loadings to be unity, and requiring the unique variances to be identical. Under this model, the covariance structure in Equation 5 holds, and the observed variables have identical unconditional expectations

$$E(\mathbf{X}) = \mathbf{1}\kappa, \tag{6}$$

with κ a scalar. This model contains three parameters: the common factor variance, common factor mean, and the unique factor variance. The degrees of freedom for this model are therefore calculated by setting $t = 3$ in Equation 4. The constraints given in Equation 6 provide a direct test of the parallel model assumption of identical true scores.

For some purposes, a weaker version of the parallel model may be useful. This weaker model retains the assumption of identical error variances across measures, but allows true scores to differ by additive constants. Specifically, if t_i is the true score variable for the ith measure, then

$$t_i = t_j + a_{ij} \tag{7}$$

for all $i, j = 1, ..., p, i \neq j$, with the a_{ij} being unknown constants. True score variances remain identical across measures in this model. This weaker model will be denoted the *essentially parallel model*, drawing an analogy from the *essentially tau-equivalent* model discussed in Lord and Novick (1968).

In the CFA with latent means, the essentially parallel model is specified by retaining the covariance structure in Equation 5, but modifying the mean structure in Equation 6

$$E(\mathbf{X}) = \nu + \mathbf{1}\kappa \tag{8}$$

where ν is a $p \times 1$ vector of intercepts. One of the p intercepts can be fixed for identification. If the jth intercept is fixed at zero for this purpose, the latent mean κ will be estimated as the observed mean of the jth variable. The identification chooses an origin for the common factor variable. Any observed variable may be chosen for this purpose, as the choice does not affect the fit of the model or the estimates of the common and unique factor variances. There are $p + 2$ parameters to be estimated in the essentially parallel model. The degrees of freedom for this model are calculated by setting $t = p + 2$ in Equation 4.

The parallel model is a special case of the essentially parallel model in which all intercepts are fixed at zero. For some purposes, only essential parallelism is required. For example, it can be shown that the Spearman-Brown formula holds under essential parallelism as well as strict parallelism. The correlation between two essentially parallel measures gives the reliability of those measures, as is true for strictly parallel measures. Furthermore, the essentially parallel model may provide a more realistic description of actual data. Observed measures which are similar in content but differ in difficulty may be more adequately fit by the essentially parallel model.

Tau-equivalent Models

Tau-equivalent measures have identical true scores but may have unequal error variances. In a CFA with zero latent means, tau-equivalence is created by requiring all measures to have equal (unit) factor loadings, while allowing the unique variances to vary (Jöreskog, 1971). These restrictions give the covariance structure

$$\Sigma = \phi \mathbf{1}\mathbf{1}' + \theta \tag{9}$$

with θ a $p \times p$ diagonal matrix. This model contains $p + 1$ parameters to be estimated.

In a CFA with nonzero latent means, we can distinguish between tau-equivalence and essential tau-equivalence (Lord & Novick, 1968). Tau-equivalent measures have identical true scores, whereas essentially tau-equivalent measures have true scores which differ by additive constants as in Equation 7. Error variances may differ in both models, but true score variances are identical. The tau-equivalent model is implemented by combining the covariance structure in Equation 9 with the mean structure in Equation 6. The common factor mean, common factor variance, and the p unique factor variances will each be estimated, a total of $p + 2$ parameters. The degrees of freedom for this model are found by setting $t = p + 2$ in Equation 4. The essentially tau-equivalent model is created by combining the covariance structure in Equation 9 with the mean structure in Equation 8. The intercept identification problem is the same as that in the essentially parallel model and can be resolved in the same way. The essentially tau-equivalent model requires $2p + 1$ parameters, and the degrees of freedom are calculated by setting $t = 2p + 1$ in Equation 4.

Essential tau-equivalence is a weaker condition than tau-equivalence, but is sufficient for some applications. As an example, the alpha coefficient provides an exact reliability estimate, rather than a lower bound, under either tau-equivalence or essential tau-equivalence (Lord & Novick, 1968). In ability or achievement testing applications, essentially tau-equivalent measures vary in difficulty but measure *the same thing*, whereas tau-equivalent measures are of equal difficulty. Both tau-equivalent models allow the observed measures to differ in reliability, unlike the parallel models.

Congeneric Models

Congeneric measures, first proposed by Jöreskog (1971) and anticipated by Meredith (1965), have linearly equivalent true scores and unequal error variances. Linear equivalence holds if the true scores on any given measure are a linear function of the true scores on any other measure:

$$t_i = B_{ij}t_j + a_{ij} \tag{10}$$

for $i, j = 1, ..., p, i \neq j$, with B_{ij} and a_{ij} being unknown constants. As discussed by Jöreskog (1971), congeneric measures share a single common factor. Equation 10 implies that true score means and variances may differ across congeneric measures, allowing different true score scales. This flexibility is often desirable in models for psychological measurements. If Equation 10 does not hold, the true scores on the *i*th and *j*th measures will not be perfectly correlated, and the measures do not share a common factor. In this case, we might conclude that the two measures are no longer measuring *the same thing*, although generalizations that would allow nonlinear relations among true scores could also be considered (Meredith, 1965). In a CFA with zero latent means, the congeneric model is specified by allowing both the factor loadings and the unique variances to vary across measures, giving a covariance structure

$$\Sigma = \phi\lambda\lambda' + \theta \tag{11}$$

with λ a $p \times 1$ vector of factor loadings. To identify the model, one factor loading can be fixed to a nonzero value (usually unity) or the common factor variance can be fixed, leaving $2p$ parameters to be estimated.

In a CFA with nonzero latent means, we can distinguish between congeneric and *essentially congeneric* models. For consistency with the previously defined essentially parallel and tau-equivalent models, we will describe measures as *essentially congeneric* if Equation 10 holds with nonzero a_{ij}, and *congeneric* if $a_{ij} = 0$ for all i and j in Equation 10. Congeneric measures have true scores that are strictly proportional. The congeneric model can be specified in a CFA with latent means by combining the covariance structure in Equation 11 with the mean structure

$$E(X) = \Lambda\kappa. \quad (12)$$

This model can be identified by fixing one factor loading to unity, or by fixing the common factor variance to unity, leaving $2p + 1$ parameters to be estimated. The degrees of freedom for this model are found by setting $t = 2p + 1$ in Equation 4.

In a CFA with nonzero latent means, the essentially congeneric model is specified by combining the covariance structure in Equation 11 with the mean structure in Equation 3. One intercept can be fixed for identification. If one of the factor loadings has been fixed for identification, we can fix the intercept corresponding to the variable whose factor loading is fixed. A convenient value for the fixed intercept is zero. If the corresponding factor loading is fixed at unity, the common factor mean κ is estimated as the mean of the observed variable whose loading is fixed. There are $3p$ parameters to be estimated in this model, and the degrees of freedom are calculated by setting $t = 3p$ in Equation 4.

If the congeneric model holds, differences in observed means are attributable to differences in the factor loadings among the measures. Furthermore, any constraints that are placed upon these loadings will alter the fit of the model to the observed mean structure. If the essentially congeneric model holds, differences in the observed means may be due to different intercepts, different factor loadings, or both. Constraints placed upon the factor loadings may not directly affect the fit of the model to the observed mean structure because changes in the loadings could be offset by shifts in the intercepts.

To summarize, we have considered six different measurement models which may be distinguished using CFA with nonzero latent means. The models have different implications for observed moment structures: means, variances, and covariances. For example, only the parallel and tau-equivalent models imply equal observed means, and only the parallel and essentially parallel models imply equal observed variances. All of the models except the congeneric and essentially congeneric models imply equal covariances among all pairs of observed measures. The parallel and essentially parallel models imply equal correlations among such pairs.

Evaluation of Fit

The logical nesting among the six measurement models can provide a basis for hierarchical fit-testing. All of the models can be viewed as special cases of the essentially congeneric model, with the parallel model being the most restrictive. The parallel, tau-equivalent, and congeneric models are each nested within their *essential* counterparts. The congeneric and essentially tau-equivalent models are not nested, and cannot be hierarchically compared. The same condition holds for the tau-equivalent and essentially parallel models.

Although the traditional chi-square statistic can be used for the evaluation of fit, most researchers will use additional goodness-of-fit indices. Fit indices in structural equation models have been intensively studied in recent years (Akaike, 1987; Bentler & Bonett, 1980; Cudeck & Browne, 1983; Hoelter, 1983; James, Mulaik, & Brett, 1982; La Du & Tanaka, 1989; Marsh, Balla, & McDonald, 1988; Mulaik, James, Van Alstine, Bennett, Lind, & Stilwell, 1989; Sobel & Bohrnstedt, 1985; Tanaka, 1987; Tanaka & Huba, 1985). Several fit indices, such as the normed fit index (Bentler & Bonett, 1980) or the parsimonious fit index (James, Mulaik, & Brett, 1982), require specification of a null model whose fit serves as a reference point for comparisons. More recent fit indices that employ estimates of the chi-square noncentrality parameter (Bentler, 1990; McDonald & Marsh, 1990) may also assess comparative fit in relation to the null model. The typical null model in the CFA with zero latent means assumes no common factors implying an observed covariance matrix that is diagonal. In the CFA with nonzero latent means, a null model that assumes no common factors can also be created, but additional restrictions might be considered. Two of the measurement models described earlier entail equality restrictions on the unique variances. These restrictions can be incorporated in the null model as well. In addition, the mean structure can be restricted in the null model in various ways. One option would add restrictions that imply equal observed means, without specifying their common value. A useful null model that incorporates both variance and mean equality restrictions can be implemented by restricting the common factor variance to be zero in the parallel model described earlier. The resulting null model contains two parameters: the unique factor variance and the latent mean. If this model is implemented through the analysis of the observed moment matrix, the model does not imply a diagonal moment matrix.

It should be noted that the *parallel* model as traditionally represented in the CFA with zero latent means is equivalent to the essentially parallel model in the nonzero latent means CFA. The chi-square fit statistic and degrees of freedom are the same for these two models. Similarly, the *tau-equivalent* model as implemented in the CFA with zero latent means is equivalent to the essentially tau-equivalent model in the nonzero latent means CFA, and the *congeneric* model in the CFA with zero latent means is equivalent to the essentially congeneric model described earlier.

Uniqueness and Error

One difficulty in using CFA to test measurement models implied by classical test theory concerns the relationship between the unique factor scores in factor analysis and the error scores in test theory. By assumption, these two sets of scores have similar statistical properties. But traditionally, the unique factor variance may include both the error score variance and a portion of the true score variance, the specific variance. Conceptually, the specific variance is that portion of the true score variance that is not linearly related to the postulated common factor(s). To illustrate, suppose that the factor analysis could be conducted using true scores, rather than observed scores. The specific variance would be equal to the unique variance in such an analysis. Alternatively, we can view the specific variance as the residual variance in the regression of the true scores on the common factors. In a confirmatory analysis the specific variance will depend upon the number of common factors and their hypothesized structure. The choice of which observed variables are included in the set to be analyzed should also affect the magnitude of the specific variance.

The specific variance is not separately identified as a model parameter and cannot be estimated in the standard common factor model. Two alternative approaches can be used to check for the existence of specific variance. If accurate reliability estimates are available, the error variance can be directly estimated and compared with the unique variance estimate provided by the factor analysis. Alternatively, second-order factor analysis can be used to separate the specific and error portions of the unique variance. Jöreskog (1971), and Rindskopf and Rose (1988) describe the use of second-order factor analysis for this purpose. The specific variance is estimated as the unique variance in the second-order factor analysis. The use of second-order factor analysis requires multiple first-order factors. Rindskopf and Rose (1988) review the conditions which must be met for identification in a second-order analysis.

If no estimation of the specific variance is possible, the assumption of no specific variance should be made only after careful consideration of the observed variables and the postulated common factors. The specific variance is zero only if the common factors account for all true score variation in the observed measures. If the specific variance is nonzero, equality of unique variances is not equivalent to equality of error variances across observed measures. Parallel measures cannot be distinguished from tau-equivalent measures using CFA in this case. If we remove the parallel/tau-equivalent distinction and label all such measures tau-equivalent, we are left with four measurement models: tau-equivalent, essentially tau-equivalent, congeneric and essentially congeneric. These four models may be distinguished within a CFA using nonzero latent means, allowing a direct empirical check on the assumption of equal true scores in the tau-equivalent model.

Examples

The first example illustrates each of the six measurement models in data provided by 200 male college freshman on the three subtests of the Reading Comprehension Test, part of the Descriptive Tests of Language Skills (College

Board, 1978). The subtests measure three components of reading comprehension: recognizing main ideas, understanding direct statements, and drawing inferences. Each subtest contains items in a multiple choice format. The subtests differ in length, and we would not expect the parallel or tau-equivalent models to provide adequate fits. The analyses reported in this example, and all subsequent examples, were performed using the LISREL VI program (Jöreskog & Sörbom, 1985). The LISREL VI program manual describes the implementation of latent means within the program. Further discussion can be found in Hayduk (1987) or Bollen (1989). Table 1 gives the augmented moment matrix used in the analysis. The row labelled *unit* gives the observed means for the three subtests. Given three observed variables, all of the measurement models are overidentified except the essentially congeneric model, which is just identified.

The six models are each fit to the data for purposes of illustration. Table 2 gives the parameter estimates and test statistics for all models. Note that in all *essential* models, the common factor mean κ is estimated as the observed mean of the first subtest because of the choice of identification in the intercepts. In the essentially parallel and essentially tau-equivalent models, the intercepts are estimated as deviations of the observed means for each subtest from the mean of the subtest whose intercept was fixed to zero for identification. This is not true in the essentially congeneric model because of the unequal factor loadings. The essentially parallel model provides an adequate and parsimonious fit to these data. The parallel and tau-equivalent models, which imply equal observed means across measures, give poor fits as expected. If the traditional *parallel*, *tau-equivalent*, or *congeneric* models are fit to these data using CFA with zero latent means, the resulting chi-square fit statistics are equal to those given by the essentially parallel, essentially tau-equivalent, or essentially congeneric models (respectively) in Table 2.

Table 1
Reading subtest moment matrix for males (N = 200)

	R1*	R2	R3	UNIT
R1	107.195			
R2	82.680	67.005		
R3	106.545	83.905	110.325	
UNIT	9.955	7.775	10.105	1.000

* R1 = Recognizing main ideas, R2 = Understanding direct statements, R3 = Drawing inferences, UNIT = Unit vector

Table 2
Parameter Estimates and Test Statistics for the Reading Subtest Data in the Male Sample

	Model[1]					
Parameter	P	EP	TE	ETE	C	EC
ν_1	0*	0*	0*	0*	0*	0*
ν_2	0*	-2.180	0*	-2.180	0*	-1.158
ν_3	0*	0.150	0*	0.150	0*	0.039
λ_1	1*	1*	1*	1*	1*	1*
λ_2	1*	1*	1*	1*	0.786	0.897
λ_3	1*	1*	1*	1*	1.015	1.011
κ	9.278	9.955	9.677	9.955	9.940	9.955
ϕ	4.956	5.523	5.596	5.430	6.275	5.884
θ_{11}	3.798	2.098	2.167	2.347	2.117	2.209
θ_{22}	3.798	2.098	6.425	1.620	2.059	1.817
θ_{33}	3.798	2.098	2.547	2.375	2.082	2.198
df	6	4	4	2	2	0
χ^2	242.88	6.59	216.46	3.47	4.43	0

[1] P = Parallel, EP = Essentially parallel, TE = Tau-equivalent, ETE = Essentially Tau-equivalent, C = Congeneric, EC = Essentially congeneric
*Fixed parameter

The second example considers the use of the measurement models in data from multiple groups of examinees. In multiple group data, we can examine the measurement properties of the variables both within and between groups. For example, we might investigate whether underlying model parameters, such as factor loadings and intercepts, are invariant across groups. Within a single group, the regression of the observed variables **X** on the common factor ξ can be expressed under the essentially congeneric model as

$$E(\mathbf{X}|\xi) = \nu + \Lambda\xi. \tag{13}$$

If the factor loadings or intercepts differ across groups, a given common factor score ξ will correspond to different expected observed scores in different groups. As a result, the factor may not have an interpretation that is common to all groups, particularly if the factor loadings show group differences. Interpretation is greatly simplified if the factor loadings and intercepts are invariant across groups. Group comparisons of latent means then have clear implications for differences in observed means. Comparisons of latent means may be particularly interesting if one or more groups have been given experimental treatments prior to measurement.

The data for the second example combine the Reading Comprehension Test data from the first example with additional data on the same variables provided by 182 female freshmen, resulting in two groups of examinees defined by gender. Table 3 gives the augmented moment matrix for the females. We first determine which measurement model provides an acceptable fit to the data within both groups. The parallel and tau-equivalent models are not considered because these models would predict equal observed means among the three measures. The essentially congeneric model is just-identified unless further invariance restrictions are imposed. The essentially parallel, essentially tau-equivalent, and congeneric models are each fit to the data. Goodness-of-fit statistics for these tests, and those which follow, are given in Table 4. The congeneric model appears to give the best fit of the three models. Although the essentially tau-equivalent model fit well in the male group in the first example, modification indices suggest that in the female group the requirement of equal factor loadings is too stringent. The congeneric model relaxes the requirement of equal factor loadings, but assumes null intercepts.

Given a measurement model that provides an adequate fit in each group, we can examine invariance restrictions on the model parameters. Within the congeneric model, we first require each variable to have the same factor loading across groups, but allow the loadings to differ among variables. The chi-square statistic given in Table 4 indicates that these invariance restrictions do not impair the fit of the model. We conclude that the regression function given in Equation 13 is identical across groups, and that the subtests provide equivalent measures across groups in this sense. As a further restriction, we require the latent means κ to be identical across groups. The chi-square statistic in Table 4 suggests that this restriction is too stringent. There appear to be gender differences in the latent

Table 3
Reading Subtest Moment Matrix for Females (N = 182)

	R1*	R2	R3	UNIT
R1	84.687			
R2	64.363	52.885		
R3	86.335	67.440	92.605	
UNIT	8.676	6.841	9.143	1.000

* R1 = Recognizing main ideas, R2 = Understanding direct statements, R3 = Drawing inferences, UNIT = Unit vector

Table 4
Test Statistics for Models in Male and Female Reading Data

Model	χ^2	*df*	χ^2 Diff	Diff *df*
Essentially Parallel	30.43	8		
Essentially Tau-equivilent	26.71	4		
Congeneric	6.41	4		
$\Lambda_m = \Lambda_f$	9.57	6	3.16	2
$\kappa_m = \kappa_f$	25.54	7	15.97	1

means. Under the previous model with invariant factor loadings, the male latent mean was estimated as 9.890 and the female mean as 8.775. The gender difference is consistent with the differences in observed means, with males receiving higher scores.

In comparing latent mean and intercept values across groups, it should be remembered that the identification chosen for these parameters will affect their estimates. If the parameters are identified by fixing a particular intercept to zero, then the same observed variable should be chosen for this purpose in each group. Suppose that essentially congeneric models are fit, with identification achieved by fixing a specific intercept to zero in each group. Without further constraints, this identification results in a latent mean estimate equal to the observed mean for the variable whose intercept was fixed in each group. Suppose that additional constraints are introduced to achieve invariance in factor loadings and intercepts across groups, as in the second example above. If a test of equality of latent means across groups is then performed, the results of this test do not depend on which intercept was fixed in the initial identification. The same chi-square statistic (and other fit indicators) will be found regardless of the initial choice of intercept. However, the values of the latent mean estimates do depend on the choice of identification. This fact must be considered in reporting the results of such an analysis.

As a final example, we consider the application of the measurement models in longitudinal data. In such data, the measurement properties of the observed variables may change over time. For example, the intercepts and factor loadings in Equation 13 may shift over time. If the measurement model remains stable or stationary over time, we can examine changes in the latent means. Changes in these means may be difficult to interpret if the measurement model is also changing. To illustrate, let $\mathbf{X}_1$ be the vector of observed variables at occasion one, and $\mathbf{X}_2$ be the vector of observed variables at occasion two, with $\mathbf{D} = \mathbf{X}_2 - \mathbf{X}_1$. Then from Equation 3, we can express the unconditional expectation of $\mathbf{D}$ as

(14) $$E(\mathbf{D}) = (\nu_2 - \nu_1) + (\Lambda_2\kappa_2 - \Lambda_1\kappa_1).$$

Clearly, equality of the latent means is not sufficient for this expectation to be zero unless stationarity in the intercepts and factor loadings also holds. If the factor loadings are stationary, then changes in the latent means are linearly related to changes in the observed means

(15) $$E(\mathbf{D}) = (\nu_2 - \nu_1) + \Lambda(\kappa_2 - \kappa_1).$$

If the further requirement of stationarity in intercepts is met, the hypothesis of no *true* mean change ($\kappa_2 = \kappa_1$) is equivalent to the hypothesis of no observed mean change: $E(\mathbf{D}) = 0$.

The data for this example are taken from a study by Nesselroade and Baltes (1974) that used measures from the Primary Mental Abilities Test (PMA) (Thurstone & Thurstone, 1962). The examinees are 99 female students measured on two occasions during grades 7 and 8. We only present data on three of the six subtests from the PMA: verbal meaning, number facility, and word groupings. These three subtests can be viewed as measures of crystallized intelligence (although number facility may have a fluid component), and therefore it is reasonable to attempt to fit a single common factor to them. The augmented moment matrix is given in Table 5. We begin the analysis with a model that assumes the measures are essentially congeneric within each occasion, with no stationarity restrictions:

(16) $$\begin{vmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{vmatrix} = \begin{vmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \\ \nu_6 \end{vmatrix} + \begin{vmatrix} 1.0 & 0 \\ \lambda_2 & 0 \\ \lambda_3 & 0 \\ 0 & 1.0 \\ 0 & \lambda_5 \\ 0 & \lambda_6 \end{vmatrix} \begin{vmatrix} \xi_1 \\ \xi_2 \end{vmatrix} + \begin{vmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{vmatrix}$$

In the above, x_1 to x_3 are the three subtests measured on the first occasion, and x_4 to x_6 are the three subtests measured on the second occasion. Note that there are two factors, one for each occasion, and that each variable loads on only a single factor. The two common factors may freely correlate across occasions. In longitudinal data, it may be unreasonable to assume that the unique factors are mutually uncorrelated. We begin by allowing autocorrelations between unique factors corresponding to the same variable measured on two occasions. For example, δ_1 and δ_4 may be correlated.

Table 6 gives goodness-of-fit statistics for the models fit to these data. The *baseline* model is essentially congeneric as described above, and this model gives

Table 5
Moment matrix for PMA Data ($N = 99$)

	Verbm1	Numbf1	Wordg1	Verbm2	Numbf2	Wordg2	Unit
Verbm1	80.586						
Numbf1	88.141	121.232					
Wordg1	114.354	143.758	194.535				
Verbm2	97.303	115.202	150.515	133.152			
Numbf2	109.121	141.899	178.606	141.980	184.818		
Wordg2	127.061	158.172	207.838	165.960	196.929	234.495	
Unit	8.000	10.202	13.485	10.505	12.657	14.838	1.000

Note. Key to abbreviations: Verbm = Verbal Memory, Numbf = Number Facility, Wordg = Word Grouping

a good fit. The next model imposes stationarity on the factor loadings. This constraint does not impair the fit of the model. The following model adds the constraint of stationary intercepts. As indicated by the difference in the chi-square statistics, this additional constraint fits marginally ($p = .10$), but we will retain the constraint. The next model equates the factor loadings for the three observed measures, setting the common value to unity. This constraint changes the measurement model to be essentially tau-equivalent within each occasion, and gives an adequate fit. Although the factor loadings are now equal and stationary, the common factor variance is not constrained to stationarity. The next model attempts to equate all intercepts to zero. This constraint worsens the fit of the model, and is not retained. Finally, the latent mean is constrained to be stationary, implying no true growth across occasions. This constraint also worsens the fit of

Table 6
Test statistics for models in the PMA data

Model	χ^2	*df*	χ^2 Diff	Diff *df*
Baseline	1.48	5		
Stationary Loadings	1.69	7	.21	2
Stationary Intercepts	6.32	9	4.63	2
Equal Loadings	8.74	11	2.42	2
Null Intercepts	142.84	13	134.10	2
Stationary Means	73.30	12	64.56	1

the model. We conclude that while the factor loadings and intercepts are stationary, the latent mean changes across occasions. At the first occasion, this mean is estimated as 8.182 and at the second occasion the estimate is 10.244.

Discussion

We have described the use of CFA with nonzero latent means in evaluating six measurement models derived from classical test theory. The relationships among the six models, and their implications for observed moment structures, have also been described. Three examples of the use of the models in real data were presented, but additional applications exist. For example, in multivariate experimental or quasi-experimental research, any of these models could be embedded within a larger structural model to reveal treatment effects. Sörbom (1978, 1982) has described the use of latent means in the analysis of covariance as applied in experimental or quasi-experimental designs. In this application, separate measurement models are specified for the covariates and the dependent measures. Several advantages accrue from the use of explicit measurement models in this analysis. Possible biasing effects due to fallible covariates (Cochran, 1968; Lord, 1960) are removed or minimized, and the power to detect treatment effects can be increased through the removal of measurement error in the dependent measures. In some applied problems, the influence of the experimental manipulation on the psychometric properties of the dependent measures may be of interest. Millsap and Hartog (1988) used a latent means analysis to detect change in the psychometric properties of observed measures in the nonequivalent control group design and to link this change to the manipulation. In other applications, the latent means analysis can reveal group differences in the psychometric properties of the measured variables that exist prior to any intervention. These group differences can bias the results of group comparisons (Bejar, 1980).

The measurement models described earlier make no distributional assumptions concerning true, error, or observed scores. Ideally, estimation and testing under these models should employ procedures that do not require such assumptions. Maximum likelihood estimation procedures do require assumptions about the population distribution of the observed variables, or alternatively, about the distributions of the common and unique factor variables. Either the common and unique factor variables are assumed to have multivariate normal distributions, implying normality of the observed variables, or multivariate normality for the observed variables is directly assumed (Anderson & Rubin, 1956; Jöreskog, 1967; Lawley & Maxwell, 1971). Because these distributional assumptions are not really part of the measurement models themselves, maximum likelihood procedures are not ideal in the above sense. In nonnormal data, maximum likelihood procedures can provide fairly accurate parameter estimates in CFA, but may not give accurate standard errors or test statistics (Boomsma, 1983).

In large samples, generalized least squares estimators have nearly the same properties as maximum likelihood estimators, but do not require strong distributional assumptions (Browne, 1974; Jöreskog & Goldberger, 1972). Generalized least squares estimation is currently available as an option in LISREL. Recent research on asymptotically efficient estimation in structural equation models (Bentler, 1983; Browne, 1982, 1984) may lead to practical procedures that give accurate estimates and standard errors with minimal distributional assumptions. Some of these methods are incorporated in the latest version of LISREL (Jöreskog & Sörbom, 1987), in the LISCOMP program (Muthén, 1987), and in the EQS program (Bentler, 1985). These new procedures may allow efficient estimation and testing of measurement models in large samples without the need for extra distributional assumptions.

The extension of traditional CFA to allow nonzero latent means provides a flexible and powerful tool for the verification of classical measurement assumptions. But the choice of an appropriate measurement model should be based on substantive considerations in addition to the results of statistical tests. The content of the observed measures may suggest a preliminary choice, which can then be tested and modified if necessary. In other cases, the choice of the measurement model will follow from the hypothesized relationships among the variables under study or from the design of the data collection. The number of statistical tests of fit should be kept to a minimum. Confirmatory statistical analyses should be used to support, rather than dictate, the choice of a measurement model.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317-322.

Alwin, D. F., & Jackson, D. J. (1980). Measurement models for response errors in surveys: Issues and applications. In K. F. Schuessler (Ed.), *Sociological Methodology 1980* (pp. 68-119). San Francisco: Jossey-Bass.

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the third Berkeley symposium on mathematical statistics and probability* (Vol. V, pp. 111- 150). Berkeley: University of California Press.

Bejar, I. I. (1980). Biased assessment of program impact due to psychometric artifacts. *Psychological Bulletin, 87,* 513-524.

Bentler, P. M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika, 48,* 493-517.

Bentler, P. M. (1985). *Theory and implementation of EQS: A structural equations program.* Los Angeles: BMDP Statistical Software.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238-246.

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

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Boomsma, A. (1983). On the robustness of LISREL (maximum likelihood estimation) against small sample size and nonnormality. Unpublished Ph.D. dissertation, University of Groningen, Groningen.

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal, 8,* 1-24.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72-141). London: Cambridge University Press.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 62-83.

Cochran, W. G. (1968). Errors of measurement in statistics. *Technometrics, 10,* 637-666.

College Board (1978). *Descriptive tests of language skills*. New Jersey: Educational Testing Service.

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

Dwyer, J. H. (1983). *Statistical models for the social and behaviorial sciences*. Oxford: Oxford University Press.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9,* 139-164.

Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances.* Baltimore: Johns Hopkins University Press.

Hoelter, J. W. (1983). The analysis of covariance structures: Goodness-of-fit indices. *Sociological Methods and Research, 11,* 325-344.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data*. Beverly Hills, CA: Sage.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32,* 443-482.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36,* 109-133.

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika, 43,* 443-477.

Jöreskog, K. G., & Goldberger, A. S. (1972). Factor analysis by generalized least squares. *Psychometrika, 37,* 243-260.

Jöreskog, K. G., & Sörbom, D. (1985). *LISREL VI User's Guide*. Mooresville, IN: Scientific Software Inc.

Jöreskog, K. G., & Sörbom, D. (1987). New developments in LISREL. Paper presented at the National Symposium on Methodological Issues in Causal Modeling, University of Alabama, Tuscaloosa.

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

La Du, T. J., & Tanaka, J. S. (1989). Influence of sample size, estimation method, and model specification on goodness-of-fit assessments in structural equation models. *Journal of Applied Psychology, 74,* 625- 635.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method* (2nd ed.) New York: American Elsevier.

Loehlin, J. C. (1987). *Latent variable models: An introduction to factor path, and structural analysis*. Hillsdale, NJ: Lawrence Erlbaum.

Lord, F. M. (1960). Large-sample covariance analysis when the control variable is fallible. *Journal of the American Statistical Association 55,* 307-321.

Lord, F. M., & Novick, M. E. (1968). *Statistical theories of mental test scores*. Reading: Addison-Wesley.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391-410.

McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Lawrence Erlbaum.

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

Meredith, W. (1965). Some results based on a general stochastic model for mental tests. *Psychometrika, 30,* 419-440.

Millsap, R. E., & Hartog, S. B. (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology, 73,* 574-584.

Mulaik, S. A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105,* 430-445.

Muthén, B. O. (1987). *LISCOMP*. Mooresville IN: Scientific Software Inc.

Nesselroade, J. R., & Baltes, P. B. (1974). Adolescent personality development and historical change: 1970-1972. *Monographs of the Society for Research in Child Development* (Serial No. 154).

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behaviorial Research, 23,* 51-67.

Sobel, M. F., & Bohrnstedt, G. W. (1985). Use of null models in evaluating the fit of covariance structure models. In N. B. Tuma (Ed.), *Sociological Methodology 1985* (pp. 152-178). San Francisco: Jossey-Bass.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27,* 229-239.

Sörbom, D. (1978). An alternative to the methodology for analysis of covariance. *Psychometrika, 43,* 381-396.

Sörbom, D. (1982). Structural equation models with structured means. In K. G. Jöreskog and H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (pp. 183-195). Amsterdam: North-Holland.

Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness- of-fit in structural equation models with latent variables. *Child Development, 58,* 136-146.

Tanaka, J. S., & Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology, 38,* 197-201.

Thurstone, L. L., & Thurstone, T. G. (1962). *SRA Primary mental abilities*. Chicago: Science Research Associates.