identified. This could be accomplished by modifying the Mplus syntax in Table 7.14 with the following MODEL commands (the bolded syntax fixes the direct effects of the sex covariate on the six indicators to zero):

```
MODEL: SOCIAL BY S1-S3;
       AGORAPH BY A1-A3;
       SOCIAL ON SEX; AGORAPH ON SEX; S1-A3 ON SEX@0; SOCIAL
          WITH AGORAPH;
```

This analysis produces the following modification indices in regard to covariate–indicator direct effects:

	M.I.	E.P.C.	Std E.P.C.	StdYX E.P.C.
ON Statements				
A1 ON SEX	13.302	0.559	0.559	0.120
A2 ON SEX	10.102	0.549	0.549	0.105
A3 ON SEX	42.748	-1.046	-1.046	-0.214

As in the planned analysis, this result clearly suggests the salience of the direct effect of the sex covariate on the A3 indicator (e.g., modification index = 42.78). Because a substantive argument could be made in regard to the source of this noninvariance, this parameter could be freed in a subsequent analysis (i.e., the model depicted in the path diagram in Figure 7.5). Although the initial analysis suggests that direct effects may exist between Sex and the A1 and A2 indicators (e.g., modification indices = 13.30 and 10.10, respectively), these relationships are less interpretable and are not salient (i.e., modification indices drop below 4.0) after the direct effect of Sex → A3 is added to the model (cf. Jöreskog, 1993, and guidelines for model revision in Chapter 5).

SUMMARY

This chapter introduced the reader to the methods of evaluating the equivalence of CFA parameters within and across groups. These procedures provide a sophisticated approach to examining the measurement invariance and population heterogeneity of CFA models. Invariance evaluation may be the primary objective of an investigation (e.g., determine the generalizability of a test instrument) or may establish the suitability of subsequent analyses (e.g., verify longitudinal measurement invariance to justify group comparisons on structural parameters such as the equivalence of regres-

sive paths among latent variables). The chapter also discussed the methodology for the analysis of mean structures. Incorporating means into CFA allows for other important analyses such as the between-group equivalence of indicator intercepts (cf. differential item functioning) and between-group comparisons on latent means. While the latter is analogous to ANOVA, CFA provides a much stronger approach because the group comparisons are conducted in context of a measurement model (e.g., adjusting for measurement error and an error theory).

In Chapter 8, three new specific applications of CFA are discussed: higher-order factor models, CFA evaluation of scale reliability, and models with formative indicators. Although the specific applications and issues covered in the second half of this book are presented in separate chapters (i.e., Chapter 6 through Chapter 10), it is important to underscore the fact that these procedures can be integrated into the same analysis. For instance, the multiple-groups approach presented in this chapter can be employed to examine the between-group equivalence of a higher-order CFA, a CFA evaluation of scale reliability, or a model containing indicators that cause the latent construct (formative indicators). Missing and non-normal data are common to all types of CFA and must be dealt with appropriately (Chapter 9). These topics are covered in the next two chapters.

NOTES

1. In addition, the term "strictly parallel" has been used for indicators that, in addition to possessing invariant factor loadings and error variances, have equivalent intercepts.

2. In Figure 7.4C, the predicted value of X2 would be the same for Groups 1 and 2 at $\xi = 0$, because the intercepts are invariant. However, predicted values of X2 would be expected to differ between groups at all non-zero values of ξ.

3. The level of statistical power is the same for between-groups comparisons of parameters tested by both the MIMIC and multiple-groups approach (e.g., group differences in factor means). The difference lies in the fact that MIMIC assumes that other measurement and structural parameters (e.g., factor loadings) are equivalent across groups, and thus the measurement model is not first tested separately in each group (where sample size may not be sufficient to ensure adequate power and precision of parameter estimates within each group).

variable is either male or female. But the standardized estimate (Std = –0.261) can convey useful information about this effect. In a standardized solution, only the latent variable is standardized. Thus, the estimate of –0.261 can be interpreted as indicating that a unit increase in Sex is associated with a .261 standardized score decrease in the latent factor of Agoraphobia; or more directly, women are .261 standardized scores higher than men on the latent dimension of Agoraphobia. This value can be interpreted akin to Cohen's *d* (Cohen, 1988, 1992). Following Cohen's guidelines (d = .20, .50, and .80 for small, medium, and large effects, respectively; cf. Cohen, 1992), the sex difference for Agoraphobia is a small effect. The results in Table 7.15 also reveal that men and women do not differ with respect to Social Phobia (z = 0.69, *ns*).

Consistent with the researcher's predictions, the results of the MIMIC model show that the A3 indicator is not invariant for males and females (akin to intercept noninvariance in multiple-groups CFA). This is reflected by the significant direct effect of Sex on the A3 indicator (z = 6.65, p < .001) that is not mediated by Agoraphobia. In other words, holding the latent factor of Agoraphobia constant, there is a significant direct effect of Sex on the A3 indicator. Thus, at any given value of the latent factor, women score significantly higher on the A3 indicator than men (by .985 units, or nearly a full point on the 0–8 scale). This is evidence of *differential item functioning*; that is, the item behaves differently as an indicator of Agoraphobia in men and women. For the substantive reasons noted earlier (sex differences in the range of activities relating to personal safety), the A3 item is biased against females. Even when their level of the underlying dimension of Agoraphobia is the same as in men, women will have higher scores on the A3 indicator (cf. Figure 7.4B) because women's responses to this item ("walking alone in isolated areas") are affected by other influences that are less relevant to men (i.e., perceptions of personal safety, in addition to the underlying construct of Agoraphobia).

Although the current example had a specific hypothesis about the gender noninvariance of the social phobia-agoraphobia questionnaire (re: item A3), measurement invariance is frequently evaluated in an exploratory fashion. In the context of a MIMIC model, this can be done by fixing all direct effects between the covariate and the indicators to zero and then inspecting modification indices (and associated expected parameter change values) to determine whether salient direct effects may be present in the data. The researcher should not pursue the alternative of freely estimating all of these direct effects, because the model would be under-

TABLE 7.15. Mplus Results of MIMIC Model of Social Phobia and Agoraphobia

MODEL RESULTS

	Estimates	S.E.	Est./S.E.	Std	StdYX
SOCIAL BY					
S1	1.000	0.000	0.000	2.007	0.889
S2	1.079	0.045	23.967	2.166	0.794
S3	0.855	0.035	24.534	1.716	0.814
AGORAPH BY					
A1	1.000	0.000	0.000	1.820	0.785
A2	0.956	0.066	14.388	1.739	0.667
A3	0.917	0.063	14.495	1.669	0.684
SOCIAL ON					
SEX	-0.109	0.158	-0.690	-0.054	-0.027
AGORAPH ON					
SEX	-0.475	0.160	-2.973	-0.261	-0.130
A3 ON					
SEX	-0.985	0.148	-6.653	-0.985	-0.202
SOCIAL WITH					
AGORAPH	0.999	0.171	5.857	0.273	0.273
Residual Variances					
S1	1.072	0.126	8.533	1.072	0.210
S2	2.750	0.195	14.087	2.750	0.370
S3	1.501	0.114	13.169	1.501	0.338
A1	2.062	0.217	9.498	2.062	0.384
A2	3.777	0.264	14.301	3.777	0.555
A3	2.705	0.214	12.642	2.705	0.455
SOCIAL	4.026	0.284	14.175	0.999	0.999
AGORAPH	3.257	0.317	10.269	0.983	0.983

TABLE 7.14. *(cont.)*

Amos Basic 5.0

```
' Example of MIMIC Model in Amos 5.0

Sub Main ()
Dim sem As New AmosEngine

sem.TextOutput
sem.Standardized
sem.Smc
sem.Mods 10

sem.BeginGroup "mimic.txt"
sem.Structure "S1 = (1) SOCIAL + (1) E1"
sem.Structure "S2 = SOCIAL + (1) E2"
sem.Structure "S3 = SOCIAL + (1) E3"
sem.Structure "A1 = (1) AGORAPH + (1) E4"
sem.Structure "A2 = AGORAPH + (1) E5"
sem.Structure "A3 = AGORAPH + SEX + (1) E6"
sem.Structure "SOCIAL = SEX + (1) D1"
sem.Structure "AGORAPH = SEX + (1) D2"
sem.Structure "D1 <-> D2"

End Sub
```

the pattern matrix programming of lambda-Y (PA LX), the model specification for Social Phobia is typical (S1 is used as the marker indicator, the factor loadings of S2 and S3 are freely estimated). For Agoraphobia, A1 is used as the marker variable, and the factor loading of A2 is freely estimated. In the remaining part of PA LY (and its associated VA commands), the metric of the A3 indicator is passed onto the A3 "pseudofactor"; the metric of the sex covariate (0/1) is passed onto the Sex "factor." In the theta-epsilon pattern matrix (PA TE), the error variances of the first five indicators (S1, S2, S3, A1, A2) are freely estimated. The measurement errors of the A3 and Sex indicators are fixed to zero. Although the Sex variable will be assumed to have no measurement error in the analysis (i.e., it corresponds to known groups), the error variance of the A3 indicator will be estimated in another portion of the solution.

Because an "all Y" specification is employed, variances and covariances (not including the error variances of indicators S1 through A2) must be estimated by the psi (PS) matrix (Ψ). The first two diagonal elements of PS (i.e., ψ_{11}, ψ_{22}) freely estimate the residual variances of Social Phobia and Agoraphobia. The first off-diagonal element of PS (ψ_{21}) estimates the correlated disturbance of these latent factors. The third (ψ_{33}) diagonal element of PS estimates the residual variance of the A3 pseudofactor; this parameter estimate will in fact represent the measurement error variance of the A3 indicator. The fourth (ψ_{44}) diagonal element of PS freely estimates the variance of Sex.

The MIMIC portion of the LISREL specification resides in the pattern matrix programming of beta (BE). The beta (B) matrix has not been employed in prior examples in this book because it represents a structural component of an SEM model; in this context, "structural" refers to directional relationships among latent variables. Specifically, the beta matrix focuses on directional relationships among endogenous (Y) variables. Alternatively, the gamma matrix (Γ) focuses on the directional relationships between exogenous and endogenous variables. Because four latent Y variables have been specified, B will be a 4×4 full matrix; the row and column order of variables is SOCIAL AGORAPH A3 SEX, as specified by earlier programming. Thus, the β_{14} and β_{24} parameters in PA BE inform LISREL to freely estimate the regressive path of Sex to Social Phobia and Sex to Agoraphobia, respectively (population heterogeneity). The β_{34} element corresponds to the regressive path of Sex to A3 (measurement invariance). Finally, the β_{32} element informs LISREL to freely estimate a path from Agoraphobia to A3. Although this estimate is obtained in a structural matrix (B), it should be interpreted as the factor loading of the A3 indicator on the Agoraphobia latent factor; in fact, this estimate would equal the LY estimate if the analysis was not conducted as a MIMIC model.

The MIMIC model provides a good fit to the data, χ^2 (11) = 3.80, p = .98, SRMR = .011, RMSEA = 0.00 (90% CI = 0.00 to 0.00, CFit = 1.00), TLI = 1.008, CFI = 1.00. Inclusion of the sex covariate did not alter the factor structure or produce salient areas of strain in the solution (e.g., all modification indices < 4.0). Table 7.15 provides selected results of this solution. Of particular interest are the regressive paths linking Sex to the latent factors and the A3 indicator. As predicted, the path of Sex $\rightarrow$ Agoraphobia was statistically significant ($z = 2.97$, $p < .01$). Given how the sex covariate was coded (0 = females, 1 = males) and the negative sign of this parameter estimate (e.g., unstandardized estimate = –0.475), it can be concluded that males have a lower mean than females on the Agoraphobia factor; more specifically, the mean of females is .475 units higher than the mean of males. The completely standardized estimate of this parameter (StdYX = –0.13) is not readily interpretable because of the binary predictor (i.e., the sex covariate). In other words, it is not meaningful to discuss this relationship in terms of a standardized score change in Sex when the level of this

TABLE 7.14. Computer Syntax (Mplus, LISREL, EQS, Amos) for MIMIC Model of Social Phobia and Agoraphobia (Regressing the Latent Factors and the A3 Indicator on the Sex Covariate)

Mplus 3.11

```
TITLE:     CFA MIMIC MODEL
DATA:      FILE IS MIMIC6.DAT;
           TYPE IS STD CORR;
           NOBS ARE 730;
VARIABLE:  NAMES ARE S1 S2 S3 A1 A2 A3 SEX;
ANALYSIS:  ESTIMATOR=ML;
MODEL:     SOCIAL BY S1-S3;
           AGORAPH BY A1-A3;
           SOCIAL ON SEX; AGORAPH ON SEX; A3 ON SEX; SOCIAL WITH AGORAPH;
OUTPUT:    SAMPSTAT MODINDICES(4.00) STAND RESIDUAL;
```

LISREL 8.72

```
TITLE LISREL PROGRAM FOR MIMIC MODEL
DA NI=7 NO=730 MA=CM
LA
S1 S2 S3 A1 A2 A3 SEX                      ! SEX: 0 = FEMALE, 1 = MALE
KM
1.000
 0.705 1.000
 0.724 0.646 1.000
 0.213 0.195 0.190 1.000
 0.149 0.142 0.128 0.521 1.000
 0.155 0.162 0.135 0.557 0.479 1.000
-0.019 -0.024 -0.029 -0.110 -0.074 -0.291 1.000
SD
2.26 2.73 2.11 2.32 2.61 2.44 0.50
MO NY=7 NE=4 LY=FU,FR TE=SY,FR PS=SY,FR BE=FU,FR
LE
SOCIAL AGORAPH A3 SEX
PA LY
0 0 0 0
1 0 0 0
1 0 0 0
0 0 0 0
0 1 0 0
0 0 0 0                                    ! A3 IS A PSEUDOFACTOR
0 0 0 0
VA 1.0 LY(1,1) LY(4,2) LY(6,3) LY(7,4)    ! SET METRIC OF VARIABLES
PA TE
1
0 1
0 0 1
0 0 0 1
0 0 0 0 1
0 0 0 0 0 0
0 0 0 0 0 0 0     ! FIX ERROR OF SEX COVARIATE TO ZERO
```

(cont.)

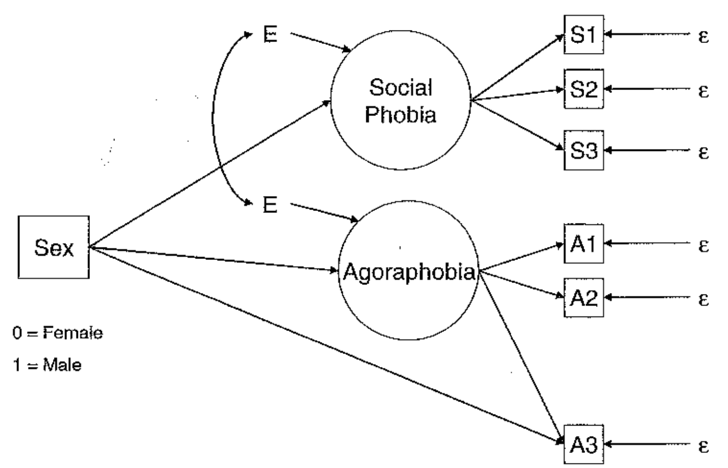
TABLE 7.14. *(cont.)*

```
PA PS
1
1 1             ! CORRELATED DISTURBANCE BETWEEN SOCIAL AND AGORAPH
0 0 1           ! ERROR VARIANCE OF A3
0 0 0 1         ! VARIANCE OF SEX
PA BE
0 0 0 1         ! PATH FROM SEX TO SOCIAL LATENT VARIABLE
0 0 0 1         ! PATH FROM SEX TO AGORAPH LATENT VARIABLE
0 1 0 1         ! PATH (FACTOR LOADING) FROM AGORAPH TO A3 INDICATOR
0 0 0 0         ! AND PATH FROM SEX TO A3 INDICATOR
OU ME=ML RS MI SC AD=OFF IT=100 ND=4
```

EQS 5.7b

```
/TITLE
 EQS SYNTAX FOR MIMIC MODEL
/SPECIFICATIONS
 CASES=730; VAR=7; ME=ML; MA=COR;
/MATRIX
1.000
 0.705 1.000
 0.724 0.646 1.000
 0.213 0.195 0.190 1.000
 0.149 0.142 0.128 0.521 1.000
 0.155 0.162 0.135 0.557 0.479 1.000
-0.019 -0.024 -0.029 -0.110 -0.074 -0.291 1.000
/STANDARD DEVIATIONS
2.26 2.73 2.11 2.32 2.61 2.44 0.50
/LABELS
 V1=s1; V2=s2; V3=s3; V4=a1; V5=a2; V6=a3; V7=sex;
 F1 = SOCIAL;
 F2 = AGORAPH;
/EQUATIONS
 V1 = F1 + E1;
 V2 = *F1 + E2;
 V3 = *F1 + E3;
 V4 = F2 + E4;
 V5 = *F2 + E5;
 V6 = *F2 + *V7 + E6;
 F1 = *V7 + D1;
 F2 = *V7 + D2;
/VARIANCES
 E1 TO E6 = *;
 D1,D2 = *;
 V7 = *;
/COVARIANCES
 D1,D2 = *;
/PRINT
 fit=all;
/LMTEST
/END
```

(cont.)



Sample Correlations and Standard Deviations (*SDs*); *N* = 730 (365 males, 365 females)

	S1	S2	S3	A1	A2	A3	Sex
S1	1.000						
S2	0.705	1.000					
S3	0.724	0.646	1.000				
A1	0.213	0.195	0.190	1.000			
A2	0.149	0.142	0.128	0.521	1.000		
A3	0.155	0.162	0.135	0.557	0.479	1.000	
Sex	-0.019	-0.024	-0.029	-0.110	-0.074	-0.291	1.000
SD:	2.260	2.730	2.110	2.320	2.610	2.440	0.500

FIGURE 7.5. MIMIC model of Social Phobia and Agoraphobia. S1, giving a speech; S2, meeting strangers; S3, talking to people; A1, going long distances from home; A2, entering a crowded mall; A3, walking alone in isolated areas (all questionnaire items rated on 0–8 scales, where 0 = no fear and 8 = extreme fear).

The first step is to ensure that the two-factor model of Social Phobia and Agoraphobia is reasonable and good fitting in the full sample (N = 730). In this step, the sex covariate is not included in the CFA and the variances (ϕ_{11}, ϕ_{22}) and covariance (ϕ_{21}) of the two factors are freely estimated (i.e., a typical CFA model is specified). This model provided a good fit to the data, $\chi^2(8) = 3.06$, $p = .93$, SRMR = .012, RMSEA = 0.00 (90% CI = 0.00 to 0.01, CFit = 1.00), TLI = 1.005, CFI = 1.00. There are no salient areas of strain in the solution (e.g., no modification indices > 4.0), and all parameter estimates were reasonable and statistically significant (e.g., range of completely standardized factor loadings = .67 to .89; correlation between Social Phobia and Agoraphobia = .28).

Next, the sex covariate is added to the model (see Figure 7.5). Table 7.14 provides Mplus, LISREL, Amos, and EQS syntax for this model specification. As can be seen in Table 7.14, the programming is straightforward in Mplus. Regressing the latent factors of Social Phobia and Agoraphobia, as well as the A3 indicator, onto the sex covariate is accomplished with the "ON" keyword (e.g., "SOCIAL ON SEX"). The correlated residual of Social Phobia and Agoraphobia is specified by the "SOCIAL WITH AGORAPH" syntax. In EQS, the MIMIC portion of the model is represented in the last three lines of the /EQUATIONS command. The line V6 = *F2 + *V7 + E6 represents the fundamental equation that variance of A3 (V6) is to be accounted for by the Agoraphobia latent factor (F2), the sex covariate (V7), and residual variance (E6). The variance of the latent factors (F1, F2) is reproduced by variance explained by the sex covariate and residual variance (e.g., F1 = *V7 + D1). The /VARIANCE commands indicate that all residual variances (indicators: E1–E6; latent factors: D1, D2) and X-variable (V7) variances should be freely estimated. The /COVARIANCE command, D1,D2 = *, indicates that the correlated disturbance (residual) between Social Phobia and Agoraphobia should also be freely estimated. Amos Basic uses similar equation (line)-based programming.

LISREL matrix programming is more complicated (see Table 7.14). Although other programs (e.g., Mplus) rely on the same programming logic exemplified by the following LISREL programming, their software contains convenience features that allow the user to set up the analysis in a less complex manner. This is also true for the SIMPLIS sublanguage of LISREL. First, note that latent-Y programming is used (lambda-Y and theta-epsilon for factor loadings and indicator errors, respectively; psi for factor variances and covariances) because the CFA portion of the MIMIC model is endogenous (cf. Chapter 3). Although Sex is really an X variable, it is treated as a Y variable so that the overall model can be specified in LISREL. An "all Y" programming strategy is often used to accomplish other special analyses in the LISREL framework (e.g., see scale reliability estimation in Chapter 8). Another programming trick is that A3 is regarded as another factor (see LE line and pattern matrix for LY); thus, the model is programmed to have four latent factors (Social Phobia, Agoraphobia, A3, Sex). This programming (in tandem with PA BE matrix programming, discussed below) actually produces a solution equivalent to a typical CFA solution (see the "Equivalent CFA Solutions" section in Chapter 5). That is, although it will not be found in the estimates for lambda-Y, the solution will generate a factor loading for A3 onto Agoraphobia that is identical to what would be obtained in an ordinary CFA specification. For on the complexity of the measurement model) because of the number of parameters that must be estimated and held equal across all groups. In addition, post hoc testing is more complex when an omnibus test of the null hypothesis of a given aspect of invariance is rejected. In contrast, MIMIC models are more parsimonious because measurement model parameters are not estimated in each group.

In most applications of MIMIC, the covariate is a nominal variable that represents levels of known groups (e.g., Sex: 0 = female, 1 = male). When the levels of group (k) are three or more, group membership can be reflected as dummy codes, per the procedures found in multiple regression textbooks (e.g., Cohen et al., 2003). If $k = 3$, two ($k - 1$) binary codes are created that identify two of the three levels of the nominal variable, and the remaining level is treated as the reference group that does not receive its own code. Although categorical covariates are typically used in MIMIC models, a normal theory estimator such as ML may still be used if the indicators of the latent factors satisfy asymptotic theory assumptions (cf. Chapter 9). This is because the categorical variables are used as predictors instead of outcomes (cf. multiple regression vs. logistic regression). However, the covariates used in MIMIC models can also be dimensional (e.g., age). The ability to accommodate continuous predictors can be viewed as another potential advantage of MIMIC over multiple-groups CFA, because the latter would necessitate imposing categorical cutoffs on a dimensional variable to form "groups" for the analysis (cf. MacCallum, Zhang, & Preacher, 2002).

Ordinarily, the covariate is assumed to be free of measurement error (i.e., its error variance is fixed to zero). This assumption is reasonable when the covariate represents known groups (e.g., male vs. female). However, the desired amount of measurement error can be modeled in a dimensional covariate using the procedures described in Chapter 4; that is, the unstandardized error of the covariate can be fixed to some non-zero value on the basis of the sample variance estimate and known reliability information (see Eqs. 4.14 and 4.15, Chapter 4).

Unlike other examples of CFA presented in this book thus far, the MIMIC approach reflects a latent-Y (endogenous) variable specification (cf. Figure 3.4, Chapter 3). This is because the latent factors (and their indicators) are specified as outcomes predicted by the covariates (see Figure 7.5). This distinction has minimal impact on syntax programming in programs such as Mplus, Amos, and EQS, which accommodate the latent Y specification "behind the scenes," but does require several alterations if the analysis is conducted using the LISREL matrix programming language.

The MIMIC methodology is illustrated using the model presented in Figure 7.5. In this example, the researcher wishes to examine selected aspects of population heterogeneity (latent factor means) and measurement invariance (indicator intercepts) associated with an established questionnaire measure of social phobia (fear of social situations due to possibility of negative evaluation by others) and agoraphobia (fear of public situations due to possibility of experiencing unexpected panic attacks; this example is loosely based on a psychometric study conducted by Brown et al., 2005). Indicators of the Social Phobia and Agoraphobia constructs are provided in Figure 7.5 (items self-rated on 0–8 scales; higher scores = higher fear). In particular, the researcher expects to find sex differences with respect to the latent construct of Agoraphobia (population heterogeneity) based on the prior clinical literature indicates that women are more inclined than men to respond to unexpected panic attacks with situational fear and avoidance. Moreover, the researcher is concerned that an indicator of Agoraphobia (A3) functions differently for men and women. Specifically, it is anticipated that, regardless of the level of the underlying factor of Agoraphobia, women will score higher on the A3 indicator than men (A3: "walking alone in isolated areas") because of sex differences in the range of activities relating to personal safety; that is, irrespective of the presence/absence of the symptoms of agoraphobia, women are less likely than men to walk alone in isolated areas. To address this question, data were collected from 730 outpatients (365 men, 365 women) who presented for assessment and treatment of various anxiety disorders.

The path diagram and input matrix of this MIMIC model are presented in Figure 7.5. A path from the covariate to the Social Phobia factor has also been specified to test for sex differences on this construct. The covariate (Sex) is represented by a single dummy code (0 = female, 1 = male), and its standard deviation and correlations with the indicators of Agoraphobia and Social Phobia are included in the input matrix. As seen in Figure 7.5, the variances of Social Phobia and Agoraphobia are not estimated in the MIMIC model. Instead, because the model attempts to explain variance in these latent factors by the sex covariate, these parameters are residual variances (E; often referred to as *disturbances*). Also note that the residual variances of Social Phobia and Agoraphobia are specified to be correlated. This specification is justified by the argument that the constructs of Social Phobia and Agoraphobia are not completely orthogonal, and that this overlap can not be fully accounted for by the sex covariate; for example, sex is not a "third variable" responsible for the overlap between the dimensions of Social Phobia and Agoraphobia.

parameters of the model (e.g., the factor loadings). For example, if the major parameters do not change when minor parameters are included in the model, this could be considered evidence for the robustness of the initial solution. The less parsimonious model may now be statistically equivalent to the baseline solution, but these additional minor parameters are not needed because they do not appreciably impact the major parameters of the solution. However, if the additional parameters significantly alter the major parameters, this would suggest that exclusion of these post hoc parameters (e.g., factor loadings or intercepts that were freely estimated in some or all groups) would lead to biased estimates of the major parameters. However, sensitivity analysis is limited somewhat by its subjective nature (i.e., what constitutes "salient" change in the model's major parameters when post hoc parameters are included or excluded from the solution). To increase the objectivity of this approach, Byrne et al. (1989) suggest correlating the major parameters (e.g., factor loadings) in the initial model with those of the best-fitting post hoc solution. Although coefficients close to 1.00 might provide clear support for the more constrained model and argue against the importance of the additional freed parameters, this conclusion becomes more vague when the correlation is well below 1.00. Moreover, a high correlation may still mask the fact that one or two major parameters differ considerably in the context of a multiple-indicator model where all other parameters are roughly the same. The caveats of partial invariance evaluation should also be considered (i.e., exploratory nature, risk of chance effects, importance of cross-validation).

MIMIC Models (CFA with Covariates)

A less commonly used method of examining invariance in multiple groups entails regressing the latent factors and indicators onto covariates that represent group membership (e.g., Sex: 0 = female, 1 = male). This approach has been referred to as *CFA with covariates* or *MIMIC models* (MIMIC = multiple indicators, multiple causes; Jöreskog & Goldberger, 1975; Muthén, 1989). Unlike multiple-groups CFA, a single input matrix is used in the analysis. The input matrix contains the variances and covariances of both the latent variable indicators and the covariates that denote group membership. Indicator means are not included as input, as might be the case for multiple-groups CFA. The two basic steps of MIMIC modeling are (1) establish a viable CFA (or E/CFA; see Chapter 5) measurement model using the full sample (i.e., collapsing across groups); and (2) add covariate(s) to the model to examine their direct effects on the latent factors

and selected indicators. A significant direct effect of the covariate on the latent factor represents population heterogeneity; that is, the factor means are different at different levels of the covariate (akin to the test of equal latent means in multiple-groups CFA). A significant direct effect of the covariate on an indicator of a latent factor represents measurement noninvariance; that is, when the latent factor is held constant, the means of the indicator are different at different levels of the covariate, thereby pointing to differential item (indicator) functioning (akin to the test of equal indicator intercepts in multiple-groups CFA). Although both aspects of invariance evaluation in MIMIC modeling correspond to the mean structure component of multiple-groups CFA, indicator intercepts and factor means are not estimated in the MIMIC analysis and indicator means are not included in the input matrix. Rather, as in the analysis of covariance structure, these group mean differences are deviations conveyed by the parameter estimates of the direct effects (e.g., Sex → Latent Factor) where the means of the indicators and factors are assumed to be zero. For instance, if the unstandardized direct effect of Sex → Latent Factor is .75, then the latent means of males and females differ by .75 units, per the metric of the factor's marker indicator.

Unlike multiple-groups CFA, MIMIC models can test only the invariance of indicator intercepts and factor means. Thus, MIMIC models assume that all other measurement and structural parameters (i.e., factor loadings, error variances/covariances, factor variances/covariances) are the same across all levels of the covariates (groups). A primary advantage of MIMIC models is that they usually have smaller sample size requirements than multiple-groups CFA. Whereas multiple-groups CFA entails the simultaneous analysis of two or more measurement models, MIMIC involves a single measurement model and input matrix. For example, in the case where a researcher wishes to conduct an invariance analysis with three groups but has a total sample size of 150, the size of each group may not be sufficient for each within-group CFA using the multiple-groups approach; that is, the analysis would require three separate CFAs, each with an $n = 50$. However, the $N = 150$ may be suitable for a single CFA in which aspects of the measurement model are regressed onto covariates, although sample size considerations must also be brought to bear with regard to the statistical power of the direct effects of the covariates (see Chapter 10).[3]

Another potential strength of MIMIC over multiple-groups CFA arises when there are many groups involved in the comparison. Multiple-groups CFA with three or more groups can be very cumbersome (depending also parameters are freely estimated (e.g., across groups) when evidence of noninvariance arises (fostered by substantive arguments, it is hoped, for why these parameters differ across groups). The final model is then fit in the second sample to determine its replicability with independent data. Although cross-validation is a compelling method to address the limitations of partial invariance evaluation (and post hoc model revision in general), Byrne et al. (1989) highlighted some practical concerns, such as the feasibility of obtaining more than one sufficiently large sample and the fact that cross-validation is likely to be unsuccessful when multiple parameters are relaxed in the first sample.

Selection of Marker Indicator

Earlier it was suggested that selection of a latent variable's marker indicator should not be taken lightly (see Chapter 4). The selection of marker indicators can also greatly influence measurement and structural invariance evaluation. Difficulties will arise in multiple-groups CFA when the researcher inadvertently selects a marker indicator that is noninvariant across groups. First, the researcher may not detect this noninvariance because the unstandardized factor loadings of the marker indicator are fixed to 1.0 in all groups. Second, subsequent tests of partial invariance may be poor fitting because the unstandardized measurement parameters of the remaining indicators (e.g., factor loadings, intercepts) are impacted by the noninvariant marker indicator. The differences found in the remaining indicators may not reflect "true" differences among groups, but rather may be an artifact of scaling the metric of the latent factor with an indicator that has a different relationship to the factor in two or more groups. Although somewhat cumbersome, one approach to exploring this possibility is to re-run the multiple-groups CFA with different marker indicators. Chueng and Rensvold (1999) have also introduced procedures for addressing this issue, although more research is needed to determine the utility and viability of these approaches.

Reliance on χ^2

As the previous data-based examples illustrate, invariance evaluation relies strongly on the χ^2 statistic. For example, the omnibus test of equality of indicator intercepts across groups is conducted by determining whether the constrained solution (in which the intercepts are held to equal across groups) produces a significant increase in χ^2 relative to a less constrained model (e.g., an equal factor loadings solution). When a significant degradation in model fit is encountered, procedures to identify noninvariant parameters rely in part on χ^2-based statistics (i.e., modification indices). As discussed in prior chapters, both model χ^2 and modification indices are sensitive to sample size. Researchers have noted that a double standard exists in the SEM literature (e.g., Chueng & Rensvold, 2000; Vandenberg & Lance, 2000). That is, given the limitations of χ^2, investigators are encouraged to use a variety of fit indices to evaluate the overall fit of a CFA solution (e.g., RMSEA, TLI, CFI, SRMR). However, in invariance evaluation, the χ^2 statistic is relied on exclusively to detect differences in more versus less constrained solutions. The reason why χ^2 is the only fit statistic used for this purpose is that its distributional properties are known and thus critical values can be determined at various degrees of freedom. This cannot be done for other fit indices; for example, a more constrained solution may produce an increase in the SRMR, but there is no way of determining at what magnitude this increase is statistically meaningful.

Researchers have begun to recognize and address this issue. Cheung and Rensvold (2000) conducted a large Monte Carlo simulation study to determine whether critical values of other goodness-of-fit statistics (e.g., the CFI) could be identified to reflect the presence/absence of measurement invariance in multiple-groups solutions; for example, does a .01- or .02-point reduction in the CFI reliably indicate the rejection of the null hypothesis that the measurement parameters are the same across groups? Although the authors proposed critical values for three fit statistics, the validity of these proposals awaits further research.

Another problem that the CFA researcher may encounter on occasion is that the omnibus test of invariance is statistically significant (i.e., the constraints result in a significant increase in model χ^2), but fit diagnostics reveal no salient strains with regard to any specific parameter; for instance, modification indices for all constrained parameters close to or below 4.0 (for an example of this outcome in the applied literature, see Campbell-Sills et al., 2004). Again, such a result may be indicative of χ^2's oversensitivity to sample size and the problem of relying exclusively on χ^2 in invariance evaluation. This outcome would suggest that the differences suggested by the statistically significant increase in model χ^2 are trivial and thus have no substantive importance. To verify this conclusion, Byrne et al. (1989) propose that a *sensitivity analysis* can be helpful. Sensitivity analysis is a post hoc method of determining whether the addition of minor parameters (e.g., relaxation of the constraints on parameters that had previously been held to equality) results in a clear change in the major unidimensional model entailing five indicators (X1–X5, where X1 is specified as the marker indicator), which is tested for invariance between two groups. Given evidence of equal form, the researcher evaluates the between-group equality of factor loadings. The more restricted solution produces a significant increase in χ^2, suggesting that at least one of the factor loadings is noninvariant. Fit diagnostics suggest that the factor loading of X5 is noninvariant (e.g., associated with high modification index; the modification indices for X2, X3, and X4 are below 4.0). This is verified in a respecified multiple-groups solution where all factor loadings other than the loading of X5 (which is freely estimated in both groups) and X1 (which was previously fixed as the marker indicator) are constrained to equality between groups; that is, this respecified model does not produce a significant increase in χ^2 relative to the equal form solution. From a statistical perspective, the invariance evaluation may proceed to examine the equality of other measurement (e.g., indicator intercepts) and structural parameters (e.g., latent means) in context of this partial invariance. The researcher would freely estimate the factor loading of X5 in both groups in subsequent analyses. Indeed, Byrne et al. (1989) note that such analyses may proceed so long as there exists at least one noninvariant parameter other than the marker indicator; for instance, in the current example, invariance evaluation could continue if, say, only the factor loading of X2 was invariant in the two groups. The same logic would apply to other measurement parameters. For example, recall that comparison of latent means is meaningful only if factor loadings and indicator intercepts are invariant. In the present illustration, a between-group comparison of the latent means could be conducted, provided that there are at least partial factor loading and partial intercept invariance; for example, at least one indicator other than the marker indicator has a invariant factor loading and intercept.

The main advantage of this strategy is that it allows the invariance evaluation to proceed after some noninvariant parameters are encountered. This is very helpful in cases where the evaluation of structural parameters is of greatest substantive interest. For instance, consider the situation where the researcher wishes to evaluate a complex longitudinal structural equation modeling involving several latent variables. Of most interest are the structural parameters of this model; that is, the paths reflecting regressions within and among latent variables over time. Before these paths can be evaluated, the researcher must establish longitudinal measurement invariance (e.g., to rule out the spurious influence of temporal change in measurement). Without knowledge of the partial measurement invariance strategy, the researcher might abandon the evaluation of structural parameters if measurement noninvariance is detected (e.g., the factor loadings of a few indicators differ over time). However, assuming a preponderance of invariant measurement parameters, the substantively more important analyses (e.g., cross-lagged effects among latent variables) can be conducted in the context of partial measurement invariance.

The primary disadvantages of partial invariance analysis are the fact that it is exploratory in nature and that it risks capitalizing on chance. These issues may be of greater concern when the purpose of the analysis is psychometric (e.g., evaluation of the measurement invariance of a psychological questionnaire in demographic subgroups). As with other scenarios, psychometric invariance evaluation is often conducted in the absence of explicit hypothesis testing. In fact, although it is counter to the traditions of the scientific method, researchers typically conduct invariance evaluation with the hope of retaining the null hypothesis; that is, H_0: all parameters are the same across groups, thereby supporting the notion that the testing instrument is unbiased, has equivalent measurement properties, and so forth. Given the number of parameters that are constrained in invariance evaluations (especially when the measurement model contains a large number of indicators), it is possible that some parameters will differ by chance. In addition, the large sample sizes used in CFA have considerable power to detect small differences as statistically significant, especially because invariance testing relies heavily on χ^2 (see the "Reliance on χ^2" section of this chapter). There are currently no methods of determining or adjusting for such issues (Byrne et al., 1989). Moreover, the guidelines and procedures for relaxing invariance constraints have not been fully developed or studied by SEM methodologists. Vandenberg and Lance (2000) recommend that the partial invariance strategy should not be employed when a large number of indicators are found to be noninvariant. Indeed, although it is statistically possible to proceed when a single indicator other than the marker indicator is invariant, this outcome should prompt the researcher to question the suitability of the measurement model for further invariance testing. Although partial invariance evaluation is a post hoc procedure, this limitation can be allayed to some degree if the researcher can provide a substantively compelling account for the source of noninvariance (e.g., see the example of a MIMIC model later in this chapter). Finally, Byrne et al. (1989) underscore the importance of cross-validation using an independent sample. For example, if the overall sample is large enough, it can be randomly divided into two subsamples. In this strategy, the first sample is used to develop a good-fitting solution where some that men and women outpatients did not differ in their average levels of the underlying dimension of Major Depression, $\chi^2_{\text{diff}}(1) = 1.92$, *ns*. It was noted earlier in this section that this constraint was essentially redundant with the significance test of the freely estimated latent mean of male outpatients ($z = 1.38$, *ns*). This is because only two groups were used in the analysis. Hence, the omnibus test, which is the equality constraint on all factor means (in the current case, $\kappa = 2$), is approximately the same as the significance test of this single parameter; in both, $df = 1$.

In instances involving more than two groups, a significant omnibus test (i.e., a significant increase in χ^2 when the means of a given latent variable are held to equality across groups) is typically followed by post hoc evaluation to determine the nature of the overall effect, along the lines of simple effects testing in ANOVA. If the analysis used three groups, three possible follow-up analyses, in which the factor means of two groups are constrained to equality at a time, could be conducted to identify which groups differed on their latent means (perhaps with a control for experiment-wise error in the multiple comparisons, such as the modified Bonferroni procedure; Jaccard & Wan, 1996). Although the similarities to ANOVA and the independent *t*-test are apparent, it should be reemphasized that the CFA-based approach to group mean comparison has several advantages over these more traditional methods. A key strength of the CFA approach is that it establishes whether the group comparisons are appropriate. Although traditional analyses simply assume this is the case, the multiple-groups CFA may reveal considerable measurement noninvariance (e.g., unequal form, a preponderance of noninvariant factor loadings and indicator intercepts) that contraindicates groups comparison on the latent factor mean. As will be discussed shortly, in some instances it may possible to proceed with such comparisons in the context of partial measurement invariance. Moreover, group comparisons have more precision and statistical power in CFA because the structural parameters (factor means, variances, covariances) have been adjusted for measurement error. Traditional tests such as ANOVA assume perfect reliability.

A decade ago, it was rare to see a CFA analysis of mean structures in the applied research literature, primarily because the syntax specification of such models was very complex in early versions of latent variable software packages. At this writing, such analyses have become more common in the literature, although the technique continues to be underutilized somewhat; for instance, many psychometric studies continue to rely exclusively on the analysis of the covariance matrix in the evaluation of psychological testing instruments. As shown in the prior example, the analysis of mean structures is straightforward in the latest releases of software programs such as LISREL, Mplus, Amos, and EQS. Given its advantages over traditional statistics (e.g., ANOVA) and its capability to evaluate other important substantive questions (e.g., psychometric issues such as differential item functioning), investigators are encouraged to incorporate mean structures in their applied CFA-based research.

Selected Issues in Single- and Multiple-Groups CFA Invariance Evaluation

In this section, three miscellaneous issues in CFA invariance evaluation are discussed. Although each of these issues can impact the conduct and interpretation of CFA invariance evaluation considerably, the extant SEM literature has provided minimal guidance on how these issues should be managed in applied research. Nevertheless, it is important that the researcher be aware of these issues, along with some tentative remedial strategies offered to date.

Partial Measurement Invariance

In their review of the applied CFA literature, Byrne et al. (1989) observed a widespread belief among researchers that when evidence of noninvariant measurement parameters (e.g., unequal factor loadings) is encountered, further testing of measurement invariance and population heterogeneity is not possible. As noted earlier, an omnibus test of invariance is conducted by placing equality constraints on a family of unstandardized parameters of the CFA model (e.g., factor loadings) and determining whether these restrictions produce a significant increase in model χ^2. If a significant increase in χ^2 is observed, then the null hypothesis is rejected (e.g., the factor loadings are not equal across groups). Fit diagnostics (e.g., modification indices, expected parameter change values) can assist the researcher in identifying the parameters that are noninvariant across groups, across time, and so on. Indeed, a significant omnibus χ^2 should not be interpreted as indicating that all parameters are noninvariant; for instance, this result might be obtained when there is a single noninvariant parameter within a complex measurement model.

Byrne et al. (1989) reminded researchers that, in many instances, invariance evaluation can proceed in the context of *partial measurement invariance*; that is, in CFA models where some but not all of the measurement parameters are equivalent. For ease of illustration, consider a

TABLE 7.13. *(cont.)*

```
/VARIANCES
 E1 TO E9= *;
 D1 = *;
/COVARIANCES
 E1,E2 = *;
/CONSTRAINTS
 (1,V2,F1)=(2,V2,F1);              ! FACTOR LOADING EQUALITY
 (1,V3,F1)=(2,V3,F1);
 (1,V4,F1)=(2,V4,F1);
 (1,V5,F1)=(2,V5,F1);
 (1,V6,F1)=(2,V6,F1);
 (1,V7,F1)=(2,V7,F1);
 (1,V8,F1)=(2,V8,F1);
 (1,V9,F1)=(2,V9,F1);
 (1,V1,V999)=(2,V1,V999);          ! INDICATOR INTERCEPT EQUALITY
 (1,V2,V999)=(2,V2,V999);
 (1,V3,V999)=(2,V3,V999);
 (1,V4,V999)=(2,V4,V999);
 (1,V5,V999)=(2,V5,V999);
 (1,V6,V999)=(2,V6,V999);
 (1,V7,V999)=(2,V7,V999);
 (1,V8,V999)=(2,V8,V999);
 (1,V9,V999)=(2,V9,V999);
 (1,E1,E1) = (2,E1,E1);            ! EQUAL INDICATOR ERRORS
 (1,E2,E2) = (2,E2,E2);
 (1,E3,E3) = (2,E3,E3);
 (1,E4,E4) = (2,E4,E4);
 (1,E5,E5) = (2,E5,E5);
 (1,E6,E6) = (2,E6,E6);
 (1,E7,E7) = (2,E7,E7);
 (1,E8,E8) = (2,E8,E8);
 (1,E9,E9) = (2,E9,E9);
 (1,D1,D1) = (2,D1,D1);            ! EQUAL FACTOR VARIANCE
 (1,F1,V999) = (2,F1,V999);        ! EQUAL LATENT MEAN
/PRINT
 fit=all;
/LMTEST
/END
```

Amos Basic 5.0

```
' Example of Full Msmt. Invariance in Amos 5.0
Sub Main ()
Dim sem As New AmosEngine
sem.TextOutput
sem.ModelMeansAndIntercepts
sem.Standardized
sem.Smc
```

(cont.)

TABLE 7.13. *(cont.)*

```
sem.BeginGroup "DepFEM.txt"
sem.GroupName "Females"
sem.Structure "m1 = (TX1) + (1) DEPRESS + (1) E1"
sem.Structure "m2 = (TX2) + (lam2) DEPRESS + (1) E2"
sem.Structure "m3 = (TX3) + (lam3) DEPRESS + (1) E3"
sem.Structure "m4 = (TX4) + (lam4) DEPRESS + (1) E4"
sem.Structure "m5 = (TX5) + (lam5) DEPRESS + (1) E5"
sem.Structure "m6 = (TX6) + (lam6) DEPRESS + (1) E6"
sem.Structure "m7 = (TX7) + (lam7) DEPRESS + (1) E7"
sem.Structure "m8 = (TX8) + (lam8) DEPRESS + (1) E8"
sem.Structure "m9 = (TX9) + (lam9) DEPRESS + (1) E9"
sem.Structure "E1 (TD1)"
sem.Structure "E2 (TD2)"
sem.Structure "E3 (TD3)"
sem.Structure "E4 (TD4)"
sem.Structure "E5 (TD5)"
sem.Structure "E6 (TD6)"
sem.Structure "E7 (TD7)"
sem.Structure "E8 (TD8)"
sem.Structure "E9 (TD9)"
sem.Structure "E1 <--> E2"
sem.Structure "DEPRESS (DEP_PHI)"
sem.Mean "DEPRESS", "0"

sem.BeginGroup "DepMALE.txt"
sem.GroupName "Males"
<same syntax under sem.GroupName "Females" command is repeated here>
```

ity in Mplus by default, and thus no additional programming is required unless the user wishes to override these defaults. Finally, the latent mean of males is held equal to females by the command [DEPRESS@0]; that is, the deviation between males' and females' latent mean is constrained to not differ significantly from zero. In EQS, all of the equality constraints are made in the /CONSTRAINTS section of the syntax. In the sections of the program corresponding to males' and females' measurement model, all parameters other than the marker indicator are freely estimated; that is, the syntax is identical for men and women. In the /CONSTRAINTS section that follows, all of these parameters are then held to equality except for the factor loading of the marker indicator and the error covariance of V1 and V2.

The equality constraint on the factor means examines whether groups differ in their levels of the underlying construct. As shown in Table 7.9, this constraint did not significantly degrade the fit of the model, indicating

TABLE 7.13. *(cont.)*

```
MO NX=9 NK=1 PH=IN LX=IN TD=SY,FR TX=IN KA=IN ! IN = INVARIANT
PA TD
1
1 1
0 0 1
0 0 0 1
0 0 0 0 1
0 0 0 0 0 1
0 0 0 0 0 0 1
0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1
EQ TD(1,1,1) TD(2,1,1)                ! EQUALITY CONTRAINTS FOR
EQ TD(1,2,2) TD(2,2,2)                ! RESIDUAL VARIANCES
EQ TD(1,3,3) TD(2,3,3)                !
EQ TD(1,4,4) TD(2,4,4)                !
EQ TD(1,5,5) TD(2,5,5)                !
EQ TD(1,6,6) TD(2,6,6)                !
EQ TD(1,7,7) TD(2,7,7)                !
EQ TD(1,8,8) TD(2,8,8)                !
EQ TD(1,9,9) TD(2,9,9)                !
OU ME=ML RS MI SC AD=OFF IT=100 ND=4
```

Mplus 3.11

```
TITLE:     MPLUS PROGRAM FOR GENDER INVARIANCE OF MAJOR DEPRESSION
DATA:      FILE IS "C:\MDDALL.DAT";
VARIABLE:  NAMES ARE SUBJ SEX M1-M9;
           USEVAR ARE M1-M9;
           GROUPING IS SEX (0=FEMALE 1=MALE);
ANALYSIS:  ESTIMATOR=ML;
           TYPE=MEANSTRUCTURE;
MODEL:     DEPRESS BY M1-M9;
           M1 WITH M2;
           M1 (1); M2 (2); M3 (3); M4 (4); M5 (5); M6 (6); ! EQUAL ERRORS
           M7 (7); M8 (8); M9 (9);                          ! EQUAL ERRORS
           DEPRESS (10);                           ! EQUAL FACTOR VARIANCE
MODEL MALE: M1 WITH M2;               ! FREELY ESTIMATE ERROR COVARIANCE
           [DEPRESS@0];                      ! CONSTRAINT ON LATENT MEAN
OUTPUT:    SAMPSTAT MODINDICES(10.00) STAND RESIDUAL;
```

EQS 5.7b

```
/TITLE
 EQS SYNTAX FOR GENDER INVARIANCE OF MAJOR DEPRESSION (FEMALES)
/SPECIFICATIONS
 CASES=375; VAR=9; ME=ML; MA=COR; ANALYSIS=MOM; GROUPS=2;
/MATRIX
<Insert correlation matrix for Females from Figure 7.3>
/MEANS
4.184 3.725 1.952 3.589 2.256 3.955 3.869 3.595 1.205
```

(cont.)

TABLE 7.13. *(cont.)*

```
/STANDARD DEVIATIONS
1.717 2.015 2.096 2.212 2.132 2.005 2.062 2.156 1.791
/LABELS
 V1=depmood; V2=anhedon; V3=weight; V4=sleep; V5=motor; V6=fatigue;
 V7=guilt; V8=concent; V9=suicide;
 F1 = DEPRESS;
/EQUATIONS
 V1 = *V999 + F1 + E1;
 V2 = *V999 + *F1 + E2;
 V3 = *V999 + *F1 + E3;
 V4 = *V999 + *F1 + E4;
 V5 = *V999 + *F1 + E5;
 V6 = *V999 + *F1 + E6;
 V7 = *V999 + *F1 + E7;
 V8 = *V999 + *F1 + E8;
 V9 = *V999 + *F1 + E9;
 F1 = *V999 + D1;                ! FACTOR MEAN FREED
/VARIANCES
 E1 TO E9= *;
 D1 = *;
/COVARIANCES
 E1,E2 = *;
/END
/TITLE
 EQS SYNTAX FOR GENDER INVARIANCE OF MAJOR DEPRESSION (MALES)
/SPECIFICATIONS
 CASES=375; VAR=9; ME=ML; MA=COR; ANALYSIS=MOM;
/MATRIX
<Insert correlation matrix for Males from Figure 7.3>
/MEANS
4.171 3.685 1.739 3.357 2.235 3.661 3.421 3.517 1.259
/STANDARD DEVIATIONS
1.598 2.018 2.094 2.232 2.108 2.113 2.286 2.174 1.788
/LABELS
 V1=depmood; V2=anhedon; V3=weight; V4=sleep; V5=motor; V6=fatigue;
 V7=guilt; V8=concent; V9=suicide;
 F1 = DEPRESS;
/EQUATIONS
 V1 = *V999 + F1 + E1;
 V2 = *V999 + *F1 + E2;
 V3 = *V999 + *F1 + E3;
 V4 = *V999 + *F1 + E4;
 V5 = *V999 + *F1 + E5;
 V6 = *V999 + *F1 + E6;
 V7 = *V999 + *F1 + E7;
 V8 = *V999 + *F1 + E8;
 V9 = *V999 + *F1 + E9;
 F1 = *V999 + D1;
```

(cont.)

test for equal factor variances best exemplifies why comparisons made by multiple-groups CFA rely on the unstandardized solution. The test of invariant factor variances would be meaningless if the metric of the factor was defined by fixing its variance to 1.0. Because the factor variance (ϕ) is strongly determined by the amount of variance in the marker indicator (i.e., ϕ = marker indicator variance multiplied by the marker indicator's squared completely standardized factor loading), the test of group equality of a factor variance can be regarded as an evaluation of whether the groups drew from similar ranges of the underlying construct to respond to the indicators of that construct. The question addressed by the test of factor variance equality often does not have clear substantive implications in applied research, although such evaluation is needed to establish the suitability of the potentially more interesting test of the invariance of factor covariances (i.e., are the latent variables more strongly related to each other in one group than another?). In the example of the measurement model of Major Depression, the factor variances were found to be equal in men and women, $\chi^2_{\text{diff}}(1) = 0.79$, *ns*. Because the example was based on a one-factor measurement model, the invariance evaluation of factor covariances was not relevant.

Table 7.13 presents the LISREL, Mplus, Amos, and EQS syntax for the fully invariant measurement model of Major Depression. In LISREL, all of the equality constraints except for the indicator residuals are made by commands on the Model line for males; that is, equal factor loadings: LX=IN; equal indicator intercepts: TX=IN; equal factor variances: PH=IN; equal latent means: KA=IN. The command TD=IN could also be employed to hold indicator error variances and covariances to equality. In this example, the covariance of the errors of M1 and M2 was allowed to vary freely in males and females to further illustrate the use of the EQ command. Because the indicator error variances, but not the M1–M2 error covariance, are to be held to equality, the EQ command is used. For instance, the statement EQ TD(1,1,1) TD(2,1,1) informs LISREL to hold the error variance of the M1 indicator (δ_{11}) in the first group (females, denoted by the first number "1" in parentheses after TD) equal to the error variance of the M1 indicator in the second group (males, denoted by the number "2" in parentheses after TD).

In Mplus, the equality constraints for indicator error variances and factor variance are placed on the solution using the numbers in parentheses that follow the names of the indicators and latent factor. This overrides the Mplus default of freely estimating these variances in both groups. As noted earlier, the factor loadings and indicator intercepts are held to equal-

TABLE 7.13. Computer Syntax (LISREL, Mplus, EQS, Amos) for Fully Invariant Multiple-Groups Model of Major Depression

LISREL 8.72

```
TITLE LISREL PROGRAM FOR INVARIANCE OF MAJOR DEPRESSION CRITERIA
    (FEMALES)
DA NG=2 NI=9 NO=375 MA=CM
LA
M1 M2 M3 M4 M5 M6 M7 M8 M9
KM
<Insert correlation matrix for Females from Figure 7.3>
ME
4.184 3.725 1.952 3.589 2.256 3.955 3.869 3.595 1.205
SD
1.717 2.015 2.096 2.212 2.132 2.005 2.062 2.156 1.791
MO NX=9 NK=1 PH=SY,FR LX=FU,FR TD=SY,FR TX=FR KA=FI
LK
DEPRESS
PA LX
0
1
1
1
1
1
1
1
1
VA 1.0 LX(1,1)                   ! SET THE METRIC OF THE LATENT VARIABLE
PA TD
1
1 1
0 0 1
0 0 0 1
0 0 0 0 1
0 0 0 0 0 1
0 0 0 0 0 0 1
0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1
OU ME=ML RS MI SC AD=OFF IT=100 ND=4
INVARIANCE OF MAJOR DEPRESSION CRITERIA (MALES)
DA NI=9 NO=375 MA=CM
LA
M1 M2 M3 M4 M5 M6 M7 M8 M9
KM
<Insert correlation matrix for Males from Figure 7.3>
ME
4.171  3.685  1.739  3.357  2.235  3.661  3.421  3.517  1.259
SD
1.598  2.018  2.094  2.232  2.108  2.113  2.286  2.174  1.788
```

(cont.)

graphically consistent with the results of the applied example presented thus far. In this graph, both the loading and the intercept of the indicator are equivalent between groups. Interpreted in isolation of factor loading invariance, the finding of intercept invariance suggests that both groups are expected to have the same observed value of the indicator (X2) when the latent factor (ξ) is zero. However, the result of factor loading invariance *and* intercept invariance can be interpreted as suggesting that, *for any given factor value*, the observed values of the indicator are expected to be statistically equivalent between groups (see dotted line in Figure 7.4A). This concept also upholds the statement made earlier in this chapter that group comparisons of the means of latent factors should be conducted only in the context of factor loading *and* indicator intercept invariance.

The remaining three graphs in Figure 7.4 show that if either the loading or intercept is noninvariant, the observed values of the indicator will differ between groups at a given level of the latent factor (see dotted lines).[2] For example, Figure 7.4B illustrates the situation of an equal factor loading but an unequal intercept. In this graph, although the indicator evidences the same relationship (regression slope) in both groups (i.e., a unit change in the factor is associated with the same amount of change in the indicator in both groups), the groups differ in the location parameter (origin) of the indicator, meaning that all predicted observed scores will differ at various levels of the latent factor. Group 1's predicted scores on the indicator will be higher than Group 2's across all levels of the "true score," suggesting that the indicator is biased. This is an example of *differential item functioning*, a term that is used to describe situations where an item yields a different mean response for the members of different groups with the same value of the underlying attribute (McDonald, 1999). An illustration of differential item functioning is presented in the "MIMIC Models" section of this chapter.

The final aspect of measurement invariance to be tested is the equality of the indicator error variances. As seen in Table 7.9, nested χ^2 evaluation indicates that the residual variances are equivalent in men and women patients, $\chi^2_{\text{diff}}(9) = 9.71$, *ns* (critical value of $\chi^2 = 16.92$, $df = 9$, $\alpha = .05$). The gain of 9 degrees of freedom corresponds to the 9 residual variances held to equality. Although this example used a real data set, this outcome is rare in applied research data. In fact, most methodologists regard equality of error variances and covariances to be an overly restrictive test that is usually not important to the endeavor of measurement invariance evaluation (e.g., Bentler, 1995; Byrne, 1998). For instance, the illustration in Figure 7.4A shows that prediction of a group equivalent observed score (X2) by the latent variable model (i.e., factor loadings, indicator intercepts) does not rely on the condition of equal indicator error variances. However, the test of equal indicator errors is relevant in situations where the researcher is interested in determining whether the reliability of an assessment measure is invariant across groups. Nonetheless, in their review of the extant measurement invariance literature, Vandenberg and Lance (2000) observed that applied researchers often mistake the test of invariant indicator errors as a test of equivalent reliabilities. These authors remind us (cf. Cole & Maxwell, 1985) that if a test of invariant indicator error variances is conducted with the intent of testing the equality of reliabilities across groups, this evaluation should be preceded by a test that establishes that the variance of the factor on which the indicators load is invariant. This is because reliability is defined by the proportion of true score variance to total variance (total variance = true score variance + error variance), where true score variance is reflected by the factor variance. This concept is developed further in Chapter 8 where an alterative approach to scale reliability evaluation is described. Ordinarily, this test should be considered as an evaluation of indicator error invariance, *not* as a test of invariant reliability, an endeavor that has minimal substantive importance in most research situations.

The remaining analyses pertain to group comparisons on the structural parameters of the CFA model (i.e., tests of population heterogeneity). As noted earlier, the viability of these comparisons rests on the evaluation of measurement invariance. In other words, it is not useful to compare groups on aspects of the latent factors (factor variances, factor covariances, latent means) without first ascertaining that the latent factors measure the same constructs in the same fashion in each group. Specifically, group comparisons on factor variances are meaningful only if the factor loadings have been found to be invariant. Comparisons of the factor covariances (in CFA models with > 1 latent factor) are meaningful if both the factor loadings and factor variances are invariant. Finally, evaluation of group equality of latent factor means rests on the condition of invariant factor loadings and indicator intercepts. The conceptual logic of these statements should be apparent upon review of the equations related to these parameters (e.g., $COV = r_{1,2}SD_1SD_2$; $SD^2 = \sigma^2$), presented in this and earlier chapters (e.g., Chapter 3).

Evaluation of the equality of a factor variance examines whether the amount of within-group variability (dispersion) of the construct differs across groups. Although crucial to all aspects of invariance evaluation, the

TABLE 7.12. *(cont.)*

Amos Basic 5.0

```
' Example of Equal Loadings and Intercepts in Amos 5.0

Sub Main ()
Dim sem As New AmosEngine

sem.TextOutput
sem.ModelMeansandIntercepts
sem.Standardized
sem.Smc

sem.BeginGroup "DepFEM.txt"
sem.GroupName "Females"
sem.Structure "m1 = (TX1) + (1) DEPRESS + (1) E1"
sem.Structure "m2 = (TX2) + (lam2) DEPRESS + (1) E2"
sem.Structure "m3 = (TX3) + (lam3) DEPRESS + (1) E3"
sem.Structure "m4 = (TX4) + (lam4) DEPRESS + (1) E4"
sem.Structure "m5 = (TX5) + (lam5) DEPRESS + (1) E5"
sem.Structure "m6 = (TX6) + (lam6) DEPRESS + (1) E6"
sem.Structure "m7 = (TX7) + (lam7) DEPRESS + (1) E7"
sem.Structure "m8 = (TX8) + (lam8) DEPRESS + (1) E8"
sem.Structure "m9 = (TX9) + (lam9) DEPRESS + (1) E9"
sem.Structure "E1 <—> E2"
sem.Mean "DEPRESS", "0"

sem.BeginGroup "DepMALE.txt"
sem.GroupName "Males"
sem.Structure "m1 = (TX1) + (1) DEPRESS + (1) E1"
sem.Structure "m2 = (TX2) + (lam2) DEPRESS + (1) E2"
sem.Structure "m3 = (TX3) + (lam3) DEPRESS + (1) E3"
sem.Structure "m4 = (TX4) + (lam4) DEPRESS + (1) E4"
sem.Structure "m5 = (TX5) + (lam5) DEPRESS + (1) E5"
sem.Structure "m6 = (TX6) + (lam6) DEPRESS + (1) E6"
sem.Structure "m7 = (TX7) + (lam7) DEPRESS + (1) E7"
sem.Structure "m8 = (TX8) + (lam8) DEPRESS + (1) E8"
sem.Structure "m9 = (TX9) + (lam9) DEPRESS + (1) E9"
sem.Structure "E1 <—> E2"
sem.Mean "DEPRESS", "Mn_DEP"

End Sub
```


indicator intercept; see Table 7.12). In females, the mean of the Depression factor (F1) is fixed to zero by the equation F1 = 0.0V999 + D1. "D1" reflects a residual variance (or *disturbance*, D) because, in fact, the analysis of mean structures requires the regression of the latent factors and indicators onto a constant (denoted in EQS as V999; Byrne, 1994). In the /VARIANCES section of the program, this disturbance variance is freely estimated in both groups (D1 = *), along with the indicator error variances (E1 TO E9 = *). In males, the only programming change is that the latent mean of Depression is freely estimated, F1 = *V999 + D1. The factor loadings and indicator intercepts are constrained to equality in the /CONSTRAINTS section of the syntax. For example, the command (1,V1,V999)= (2,V1,V999) informs EQS to hold the intercept of the V1 indicator (depressed mood) in the first group (women) equal to the intercept of the V1 indicator in the second group (men).

In Amos Basic, the mean structure component is brought into the solution with the "sem.ModelMeansandIntercepts" command. As in earlier examples (e.g., Table 7.5), parameters are held to equality by giving them the same names (e.g., TX1 for the intercept of the M1 indicator). The mean of the latent factor for females is fixed to zero by the statement, sem.Mean "DEPRESS", "0."

The equal measurement intercepts model is found to be good fitting and does not result in a significant degradation of fit relative to the equal factor loadings solution, $\chi^2_{\text{diff}}(8) = 12.47$, *ns* (see Table 7.9). The gain of 8 degrees of freedom (to a total of $df = 68$) is due to the additional 18 new elements of the input matrices (i.e., the 9 indicator means for men and women) minus the 10 mean structure parameters (9 intercepts held to equality, 1 freely estimated latent mean; for identification purposes, women served as the reference group by fixing their latent mean to zero). Because the factor loadings and indicator intercepts are invariant in men and women, comparison of the groups on the latent mean of Major Depression is interpretable. In men, the unstandardized parameter estimate for the latent mean is –.13 (not shown in the tables), indicating that, on average, men score .13 units below women on the dimension of Major Depression, a difference that is not statistically significant ($z = 1.38$). This lack of difference is upheld in a subsequent analysis of population heterogeneity (the final analysis in Table 7.9) that constrains the latent means to equality; that is, $\chi^2_{\text{diff}} = 1.92$, which is roughly the same as $z = 1.38^2$ (cf. Wald test, Chapter 4).

As mentioned earlier, Figure 7.4 graphically displays different combinations of factor loading and intercept (non)invariance. Figure 7.4A is

TABLE 7.12. *(cont.)*

Mplus 3.11

```
TITLE:        MPLUS PROGRAM FOR EQUAL LOADINGS AND INTERCEPTS OF MDD
DATA:         FILE IS "C:\MDDALL.DAT";
VARIABLE:     NAMES ARE ADIS SEX M1-M9;
              USEVAR ARE M1-M9;
              GROUPING IS SEX (0=FEMALE 1=MALE);
ANALYSIS:     ESTIMATOR=ML;
              TYPE=MEANSTRUCTURE;
MODEL:        DEPRESS BY M1-M9;
              M1 WITH M2;
MODEL MALE:   M1 WITH M2;  ! ALL MSMT PARAMETERS HELD EQUAL BY DEFAULT
OUTPUT:       SAMPSTAT MODINDICES(10.00) STAND RESIDUAL;
```

EQS 5.7b

```
/TITLE
 EQS SYNTAX FOR GENDER INVARIANCE OF MAJOR DEPRESSION (FEMALES)
/SPECIFICATIONS
 CASES=375; VAR=9; ME=ML; MA=COR; ANALYSIS=MOM; GROUPS=2;
/MATRIX
<Insert correlation matrix for Females from Figure 7.3>
/MEANS
4.184 3.725 1.952 3.589 2.256 3.955 3.869 3.595 1.205
/STANDARD DEVIATIONS
1.717 2.015 2.096 2.212 2.132 2.005 2.062 2.156 1.791
/LABELS
 V1=depmood; V2=anhedon; V3=weight; V4=sleep; V5=motor; V6=fatigue;
 V7=guilt; V8=concent; V9=suicide;
 F1 = DEPRESS;
/EQUATIONS
 V1 = *V999 + F1 + E1;
 V2 = *V999 + *F1 + E2;
 V3 = *V999 + *F1 + E3;
 V4 = *V999 + *F1 + E4;
 V5 = *V999 + *F1 + E5;
 V6 = *V999 + *F1 + E6;
 V7 = *V999 + *F1 + E7;
 V8 = *V999 + *F1 + E8;
 V9 = *V999 + *F1 + E9;
 F1 = 0.0V999 + D1;                    ! FEMALES ARE REFERENCE GROUP
/VARIANCES
 E1 TO E9= *;
 D1 = *;
/COVARIANCES
 E1,E2 = *;
/END
/TITLE
 EQS SYNTAX FOR GENDER INVARIANCE OF MAJOR DEPRESSION (MALES)
/SPECIFICATIONS
 CASES=375; VAR=9; ME=ML; MA=COR; ANALYSIS=MOM;
/MATRIX
<Insert correlation matrix for Males from Figure 7.3>
```

(cont.)

TABLE 7.12. (cont.)

```
/MEANS
4.171 3.685 1.739 3.357 2.235 3.661 3.421 3.517 1.259
/STANDARD DEVIATIONS
1.598 2.018 2.094 2.232 2.108 2.113 2.286 2.174 1.788
/LABELS
 V1=depmood; V2=anhedon; V3=weight; V4=sleep; V5=motor; V6=fatigue;
 V7=guilt; V8=concent; V9=suicide;
 F1 = DEPRESS;
/EQUATIONS
 V1 = *V999 + F1 + E1;
 V2 = *V999 + *F1 + E2;
 V3 = *V999 + *F1 + E3;
 V4 = *V999 + *F1 + E4;
 V5 = *V999 + *F1 + E5;
 V6 = *V999 + *F1 + E6;
 V7 = *V999 + *F1 + E7;
 V8 = *V999 + *F1 + E8;
 V9 = *V999 + *F1 + E9;
 F1 = *V999 + D1;               ! MALE LATENT MEAN FREELY ESTIMATED
/VARIANCES
 E1 TO E9= *;
 D1 = *;
/COVARIANCES
 E1,E2 = *;
/CONSTRAINTS
 (1,V2,F1)=(2,V2,F1);           ! FACTOR LOADING EQUALITY
 (1,V3,F1)=(2,V3,F1);
 (1,V4,F1)=(2,V4,F1);
 (1,V5,F1)=(2,V5,F1);
 (1,V6,F1)=(2,V6,F1);
 (1,V7,F1)=(2,V7,F1);
 (1,V8,F1)=(2,V8,F1);
 (1,V9,F1)=(2,V9,F1);
 (1,V1,V999)=(2,V1,V999);       ! INDICATOR INTERCEPT EQUALITY
 (1,V2,V999)=(2,V2,V999);
 (1,V3,V999)=(2,V3,V999);
 (1,V4,V999)=(2,V4,V999);
 (1,V5,V999)=(2,V5,V999);
 (1,V6,V999)=(2,V6,V999);
 (1,V7,V999)=(2,V7,V999);
 (1,V8,V999)=(2,V8,V999);
 (1,V9,V999)=(2,V9,V999);
/PRINT
 fit=all;
/LMTEST
/END
```

(cont.)

latent means are not estimated in the absolute sense but instead reflect average differences in the level of the construct across groups.

Table 7.12 presents the LISREL, Mplus, Amos, and EQS syntax for the invariant intercepts analysis (also holding the factor loadings to equality). In each program, women are used as the reference group by fixing their latent factor mean to zero and the latent mean for men is freely estimated. In the LISREL program, note that the indicator means have been added to the input matrix (under the ME heading), and tau-X (TX) and kappa (KA) matrices have been included on the Model line (MO) for the estimation of indicator intercepts and the latent factor mean, respectively. For females, the LISREL programming looks the same as a typical CFA analysis, except for the TX and KA commands. On the Model line, the command TX=FR informs LISREL that the vector of intercepts are to be freely estimated. The KA=FI command fixes the latent mean of Depression to zero in this group. The remaining salient aspects of the LISREL programming are found on the Model line for males. The factor loadings and indicator intercepts are constrained to equality between groups by the commands LX=IN and TX=IN, respectively (IN = constrain parameters to be invariant with those of the preceding group). Because indicator error variances and factor variances are not held to equality in this analysis, the TD=PS and PH=PS commands are used to inform LISREL that these parameters may be freely estimated using the same pattern matrix and starting values used in the prior group. Finally, the KA=FR command allows the latent mean of Depression to be freely estimated in males. As noted earlier, this estimate will reflect a deviation (difference) of males' latent mean relative to the latent mean of females.

Most of the aforementioned model specifications are handled by default in Mplus (Table 7.12). The syntax now includes the subcommand TYPE=MEANSTRUCTURE to inform Mplus that the analysis will entail both covariance and mean structures. Accordingly, the indicator means must be included in the input matrix, or they can be calculated in Mplus by inputting a raw data file as in the Table 7.12 example. The only syntax provided for men is "M1 with M2" to freely estimate the correlated error of these indicators (which in fact is redundant with another Mplus default of freely estimating residual variances and covariances in all groups). By default, the factor loadings and indicator intercepts are held to equality in Mplus.

In EQS, a key programming change is that the /EQUATIONS section now includes the intercept parameters, which are freely estimated in both groups (e.g., V1 = *V999 + F1 + E1, where *V999 is the freely estimated

TABLE 7.12. Computer Syntax (LISREL, Mplus, EQS, Amos) for Equal Factor Loadings and Equal Indicator Intercepts Multiple-Groups Model of Major Depression

LISREL 8.72

```
TITLE LISREL PROGRAM FOR EQUAL LOADINGS AND INTERCEPTS OF MDD (FEMALES)
DA NG=2 NI=9 NO=375 MA=CM                                    ! NOTE: NG = 2
LA
M1 M2 M3 M4 M5 M6 M7 M8 M9
KM
<Insert correlation matrix for Females from Figure 7.3>
ME
4.184  3.725  1.952  3.589  2.256  3.955  3.869  3.595  1.205
SD
1.717  2.015  2.096  2.212  2.132  2.005  2.062  2.156  1.791
MO NX=9 NK=1 PH=SY,FR LX=FU,FR TD=SY,FR TX=FR KA=FI
LK
DEPRESS
PA LX
0
1
1
1
1
1
1
1
1
VA 1.0 LX(1,1)                      ! SET THE METRIC OF THE LATENT VARIABLE
PA TD
1
1 1
0 0 1
0 0 0 1
0 0 0 0 1
0 0 0 0 0 1
0 0 0 0 0 0 1
0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1
OU ME=ML RS MI SC AD=OFF IT=100 ND=4
EQUAL LOADINGS AND INTERCEPTS OF MDD (MALES)   ! SEPARATE TITLE FOR MALES
DA NI=9 NO=375 MA=CM
LA
M1 M2 M3 M4 M5 M6 M7 M8 M9
KM
<Insert correlation matrix for Males from Figure 7.3>
ME
4.171 3.685 1.739 3.357 2.235 3.661 3.421 3.517 1.259
SD
1.598 2.018 2.094 2.232 2.108 2.113 2.286 2.174 1.788
MO NX=9 NK=1 PH=PS LX=IN TD=PS TX=IN KA=FR    ! PROGRAMMING SHORTCUTS
OU ME=ML RS MI SC AD=OFF IT=100 ND=4
```

(cont.)

was fixed to 1.0 in both groups to serve as the marker indicator, this measure is not involved in the equality constraints. Incidentally, note in Table 7.9 that the equal factor loading solution produced a slight improvement in the parsimony goodness-of-fit indices as compared with the equal form solution; for example, RMSEA = .043 versus .049 in the equal factor loading and equal form solution, respectively. This is due to the gain in degrees of freedom (60 vs. 52), coupled with the trivial change in model χ^2 (102.84 vs. 98.91) associated with reproducing the observed covariance matrices with fewer freely estimated parameters (i.e., increased model parsimony via constraint of previously free parameters to equality).

Because the constraint of equal factor loadings did not significantly degrade the fit of the solution, it can be concluded that the indicators evidence comparable relationships to the latent construct of Major Depression in men and women. Figure 7.4 graphically illustrates various forms of measurement (non)invariance with respect to factor loadings and indicator intercepts. Although the equality of intercepts has yet to be evaluated, the result of invariant factor loadings is depicted by Figures 7.4A and 7.4B, which show parallel regression slopes for Groups 1 and 2; in other words, a unit change in the underlying dimension (ξ, or Major Depression) is associated with statistically equivalent change in the observed measure (X2, or the indicator of loss of interest in usual activities) in both groups (men and women). However, because the intercepts have not been evaluated, it cannot be concluded that men and women would evidence equivalent observed *scores* on an indicator at a given level of the latent factor (as is shown in Figure 7.4A where both the loading and intercept of an indicator, X2, is equivalent between groups).

The previous analyses were based on covariance structures. To examine the between-group equality of indicator intercepts, the means of the indicators must be input to the analysis along with the indicators' variances and covariances. As in the invariance evaluation of longitudinal measures, the analysis of mean structures poses additional identification issues. In the case of this two-group analysis, there are 18 indicator means (9 for men, 9 for women) but potentially 20 freed parameters of the mean structure solution (18 intercepts, 2 latent means). Moreover, latent variables must be assigned an origin in addition to a metric. Thus, as with the longitudinal invariance example, the mean structure component of the multiple-groups solution is underidentified in the absence of additional restrictions. In addition to holding the indicator intercepts to equality across groups in the measurement invariance solution, identification can be accomplished by fixing the origin (mean) of the latent variable(s) in

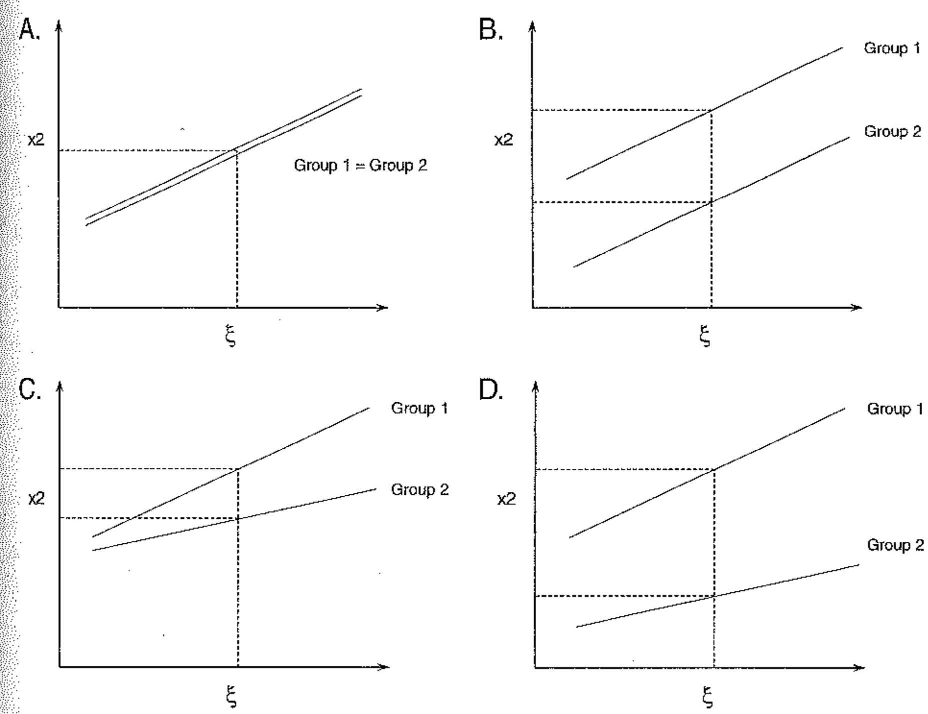


FIGURE 7.4. Graphical depictions of various forms of measurement invariance and noninvariance. (A) Equal loadings and intercepts; (B) equal loadings, unequal intercepts; (C) unequal loadings, equal intercepts; (D) unequal loadings and intercepts.

one group to zero. The group whose latent mean(s) have been fixed to zero becomes the reference group. From a statistical standpoint, selection of the reference group is arbitrary, although this choice might be guided by substantive/interpretative considerations (e.g., reference group = participants who did not receive an intervention/experimental manipulation). The latent means in the remaining groups are freely estimated, but these parameter estimates represent deviations from the reference group's latent mean. For example, if Group 2's latent M = 1.4, this is indicates that, on average, this group scores 1.4 units higher than the reference group on the latent dimension, based on the metric of the marker indicator. The rationale of this approach to model specification is that because the measurement intercepts are constrained to equality across groups (to test for intercept invariance), the latent factors have an arbitrary origin (mean). Thus,

TABLE 7.11. Parameter Estimates (Mplus) from the Equal Form Measurement Model of Major Depression in Men and Women

```
MODEL RESULTS

                      Estimates    S.E.   Est./S.E.    Std     StdYX

Group FEMALE

DEPRESS  BY
  M1                    1.000     0.000      0.000    1.250    0.729
  M2                    1.107     0.086     12.907    1.384    0.688
  M3                    0.729     0.101      7.221    0.911    0.435
  M4                    0.912     0.108      8.406    1.140    0.516
  M5                    0.812     0.104      7.845    1.016    0.477
  M6                    0.924     0.100      9.240    1.155    0.577
  M7                    0.611     0.098      6.220    0.764    0.371
  M8                    0.979     0.107      9.131    1.224    0.569
  M9                    0.484     0.085      5.707    0.606    0.339
M1  WITH
  M2                    0.394     0.147      2.688    0.394    0.114

Variances
  DEPRESS               1.563     0.224      6.991    1.000    1.000

Residual Variances
  M1                    1.376     0.155      8.856    1.376    0.468
  M2                    2.133     0.223      9.579    2.133    0.527
  M3                    3.551     0.277     12.837    3.551    0.810
  M4                    3.583     0.290     12.351    3.583    0.734
  M5                    3.501     0.278     12.609    3.501    0.772
  M6                    2.676     0.226     11.822    2.676    0.667
  M7                    3.658     0.279     13.113    3.658    0.862
  M8                    3.137     0.264     11.904    3.137    0.677
  M9                    2.831     0.214     13.223    2.831    0.885

Group MALE

DEPRESS  BY
  M1                    1.000     0.000      0.000    1.024    0.642
  M2                    1.236     0.098     12.580    1.266    0.628
  M3                    0.786     0.133      5.911    0.805    0.385
  M4                    1.166     0.152      7.656    1.193    0.535
  M5                    0.959     0.139      6.915    0.982    0.466
  M6                    1.132     0.145      7.790    1.159    0.549
  M7                    0.766     0.143      5.361    0.784    0.344
  M8                    1.019     0.144      7.075    1.043    0.480
  M9                    0.632     0.113      5.617    0.647    0.362
M1  WITH
  M2                    0.920     0.160      5.743    0.920    0.286
```

(cont.)

TABLE 7.11. *(cont.)*

```
Variances
DEPRESS                 1.048     0.183      5.719    1.000    1.000

Residual Variances
  M1                    1.499     0.152      9.888    1.499    0.588
  M2                    2.459     0.244     10.084    2.459    0.605
  M3                    3.727     0.290     12.830    3.727    0.852
  M4                    3.547     0.304     11.671    3.547    0.713
  M5                    3.467     0.282     12.304    3.467    0.783
  M6                    3.111     0.270     11.516    3.111    0.699
  M7                    4.599     0.353     13.030    4.599    0.882
  M8                    3.626     0.297     12.192    3.626    0.769
  M9                    2.770     0.214     12.943    2.770    0.869

R-SQUARE

Group FEMALE

Observed
Variable   R-Square

M1           0.532
M2           0.473
M3           0.190
M4           0.266
M5           0.228
M6           0.333
M7           0.138
M8           0.323
M9           0.115

Group MALE

Observed
Variable   R-Square
M1           0.412
M2           0.395
M3           0.148
M4           0.287
M5           0.217
M6           0.301
M7           0.118
M8           0.231
M9           0.131
```

Note. $N = 750$.

the same *pattern* and *starting* values of fixed and freed parameters of the corresponding matrix of the prior group; this is not to be confused with *keyword* "PS," which would appear to the left of the equals sign in the model specification of latent-Y variable variances and covariances. Thus, because no invariance constraints are placed in the equal form solution, the PS keyword value allows the user to inform LISREL that that the parameter specification for the measurement model for men and women are identical, without having to repeat the syntax previously written for women (i.e., pattern matrices, marker indicators).

The Mplus example in Table 7.10 runs the analysis on a raw data file instead of variance–covariance matrices (although multiple-groups CFA can also be conducted in Mplus by reading separate input matrices). In this case, the data file contains the subject identification number (SUBJ) and dummy code for sex (SEX: 0 = female, 1 = male), in addition to the nine clinical ratings of MDD (M1–M9). The USEVARIABLE command (USEVAR) selects out the nine indicators that will be used in the CFA. The GROUPING command identifies the variable in the data set that denotes the levels of group that will be used in the multiple-groups CFA. The first MODEL command is the same as would be used to specify the Figure 7.3 measurement model in a single group; in this instance, the first level of GROUP (females) identified by the GROUPING command. Recall that an Mplus default is to fix the first indicator listed to load on a latent factor as the marker indicator; thus, the M1 indicator has been automatically set to be the marker indicator. In multiple-groups analysis, Mplus holds some measurement parameters to equality across groups by default; specifically, the factor loadings as well as intercepts if indicator means are included in the model. Because of the overly stringent nature of this restriction, Mplus does not hold indicator residual variances and covariances to equality by default. Moreover, all structural parameters (factor variances, covariances, latent means) are freely estimated in all groups by default. Thus, because the current model is testing for equal form, the Mplus default for holding factor loadings to equality must be overridden. This is accomplished by the MODEL MALE: command, in which all parameters listed after this keyword are freely estimated in men's solution (however, note that the indicator list omits M1 because it was previously fixed to 1.0 to serve as the marker indicator). Note that while additional programming is necessary to freely estimate that factor loadings in men's solution (i.e., DEPRESS BY M2–M9), the line for correlated residuals between M1 and M2 (i.e., M1 with M2) is included only for clarity (it is redundant with the Mplus default of freely estimating residual variances and covariances in all groups).

As shown in Table 7.9, this solution provides an acceptable fit to the data. This solution will serve as the baseline model for subsequent tests of measurement invariance and population heterogeneity. The parameter estimates for each group are presented in Table 7.11. Inspection of Table 7.9 shows that the *df* and model χ^2 of the equal form solution equal the sum of the *df*s and model χ^2s of the CFAs run separately for men and women; for example, $\chi^2 = 98.91 = 45.96 + 53.95$. Although multiple-groups solutions can be evaluated when the size of the groups vary, if the group sizes differ markedly, interpretation of the analysis may be more complex. This is because many aspects of the CFA are influenced by (sensitive to) sample size. For instance, recall that model χ^2 is calculated as $F_{ML}(N - 1)$. Consider the scenario where the fit function value is the same in two groups (i.e., $F_{ML1} = F_{ML2}$), but the size of groups differ considerably (e.g., $n_1 = 1{,}000$, $n_2 = 500$). Thus, although the discrepancies between the observed and predicted covariance matrices are the same in both groups, the model χ^2s of the groups will differ greatly, and Group 1 will contribute considerably more to the equal form χ^2 than Group 2. Specifically, in this contrived example, Group 1 will contribute two times as much to the overall χ^2 than Group 2. All other aspects of the CFA model that are based on χ^2 (e.g., overall fit statistics such as the CFI; modification indices) or are influenced by sample size (e.g., standard errors, power to detect parameter estimates as significantly different from zero, standardized residuals) will also be differentially impacted by the unbalanced group sizes. Thus, although it is permissible to conduct multiple-groups CFA with unequal sample sizes, it is preferable for the sizes of the groups to be as balanced as possible. In instances where the group *n*s differ considerably, the researcher must be mindful of this issue when interpreting the results.

The next analysis evaluated whether the factor loadings (unstandardized) of the MDD indicators were equivalent in men and women. The test of equal factor loadings is a critical test in multiple-groups CFA. In tandem with other aspects of measurement invariance evaluation (e.g., equal form), this test determines whether the measures have the same meaning and structure for different groups of respondents. Moreover, this test establishes the suitability of other group comparisons that may be of substantive interest (e.g., group equality of factor variances, factor means, or regressive paths among latent variables). In the current data set, the equal factor loadings model had an overall good fit to the data and did not significantly degrade fit relative to the equal form solution, $\chi^2_{diff}(8) = 3.93$, *ns* (critical value of $\chi^2 = 15.51$, $df = 8$, $\alpha = .05$). The difference in degrees of freedom ($df = 8$) corresponds to the eight factor loadings (M2–M9) that were freely estimated in both groups in the previous analysis. Because M1

TABLE 7.10. *(cont.)*

```
 V3 = *F1+E3;
 V4 = *F1+E4;
 V5 = *F1+E5;
 V6 = *F1+E6;
 V7 = *F1+E7;
 V8 = *F1+E8;
 V9 = *F1+E9;
/VARIANCES
 E1 TO E9= *;
 F1 = *;
/COVARIANCES
 E1,E2 = *;
/END
/TITLE
 EQUAL FORM OF MAJOR DEPRESSION (MALES)
/SPECIFICATIONS
 CASES=375; VAR=9; ME=ML; MA=COR; ANALYSIS=COV;
/MATRIX
1.000
0.689  1.000
0.204  0.218  1.000
0.335  0.284  0.315  1.000
0.274  0.320  0.153  0.265  1.000
0.333  0.333  0.221  0.364  0.268  1.000
0.258  0.211  0.114  0.139  0.185  0.132  1.000
0.319  0.346  0.176  0.207  0.231  0.279  0.146  1.000
0.316  0.269  0.111  0.140  0.117  0.131  0.263  0.163  1.000
/STANDARD DEVIATIONS
1.598  2.018  2.094  2.232  2.108  2.113  2.286  2.174  1.788
/LABELS
 V1=depmood; V2=anhedon; V3=weight; V4=sleep; V5=motor; V6=fatigue;
 V7=guilt; V8=concent; V9=suicide;
 F1 = DEPRESS;
/EQUATIONS
 V1 = F1+E1;
 V2 = *F1+E2;
 V3 = *F1+E3;
 V4 = *F1+E4;
 V5 = *F1+E5;
 V6 = *F1+E6;
 V7 = *F1+E7;
 V8 = *F1+E8;
 V9 = *F1+E9;
/VARIANCES
 E1 TO E9= *;
 F1 = *;
/COVARIANCES
 E1,E2 = *;
/PRINT
 fit=all;
/LMTEST
/END
```

(cont.)

TABLE 7.10. *(cont.)*

Amos Basic 5.0

```
' Example of Multi-Grp Equal Form in Amos 5.0

Sub Main ()
Dim sem As New AmosEngine

sem.TextOutput
sem.Standardized
sem.Smc

sem.BeginGroup "DepFEM.txt"
sem.Structure "m1 <— DEPRESS (1)"
sem.Structure "m2 <— DEPRESS"
sem.Structure "m3 <— DEPRESS"
sem.Structure "m4 <— DEPRESS"
sem.Structure "m5 <— DEPRESS"
sem.Structure "m6 <— DEPRESS"
sem.Structure "m7 <— DEPRESS"
sem.Structure "m8 <— DEPRESS"
sem.Structure "m9 <— DEPRESS"
sem.Structure "m1 <— E1 (1)"
sem.Structure "m2 <— E2 (1)"
sem.Structure "m3 <— E3 (1)"
sem.Structure "m4 <— E4 (1)"
sem.Structure "m5 <— E5 (1)"
sem.Structure "m6 <— E6 (1)"
sem.Structure "m7 <— E7 (1)"
sem.Structure "m8 <— E8 (1)"
sem.Structure "m9 <— E9 (1)"
sem.Structure "E1 <—> E2"

sem.BeginGroup "DepMALE.txt"
sem.Structure "m1 <— DEPRESS (1)"
sem.Structure "m2 <— DEPRESS"
sem.Structure "m3 <— DEPRESS"
sem.Structure "m4 <— DEPRESS"
sem.Structure "m5 <— DEPRESS"
sem.Structure "m6 <— DEPRESS"
sem.Structure "m7 <— DEPRESS"
sem.Structure "m8 <— DEPRESS"
sem.Structure "m9 <— DEPRESS"
sem.Structure "m1 <— E1 (1)"
sem.Structure "m2 <— E2 (1)"
sem.Structure "m3 <— E3 (1)"
sem.Structure "m4 <— E4 (1)"
sem.Structure "m5 <— E5 (1)"
sem.Structure "m6 <— E6 (1)"
sem.Structure "m7 <— E7 (1)"
sem.Structure "m8 <— E8 (1)"
sem.Structure "m9 <— E9 (1)"
sem.Structure "E1 <—> E2"

End Sub
```

TABLE 7.10. Computer Syntax (LISREL, Mplus, EQS, Amos) for Equal Form Multiple-Groups Model of Major Depression

LISREL 8.72

```
TITLE LISREL PROGRAM FOR EQUAL FORM OF MAJOR DEPRESSION (FEMALES)
DA NG=2 NI=9 NO=375 MA=CM                                ! NOTE: NG = 2
LA
M1 M2 M3 M4 M5 M6 M7 M8 M9
KM
1.000
0.616  1.000
0.315  0.313  1.000
0.349  0.332  0.261  1.000
0.314  0.250  0.270  0.327  1.000
0.418  0.416  0.298  0.328  0.317  1.000
0.322  0.313  0.096  0.117  0.130  0.140  1.000
0.409  0.415  0.189  0.314  0.303  0.281  0.233  1.000
0.318  0.222  0.051  0.115  0.140  0.150  0.217  0.222  1.000
SD
1.717 2.015 2.096 2.212 2.132 2.005 2.062 2.156 1.791
MO NX=9 NK=1 PH=SY,FR LX=FU,FR TD=SY,FR
LK
DEPRESS
PA LX
0
1
1
1
1
1
1
1
1
VA 1.0 LX(1,1)                     ! SET THE METRIC OF THE LATENT VARIABLE
PA TD
1
1 1
0 0 1
0 0 0 1
0 0 0 0 1
0 0 0 0 0 1
0 0 0 0 0 0 1
0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1
OU ME=ML RS MI SC AD=OFF IT=100 ND=4
EQUAL FORM OF MAJOR DEPRESSION (MALES)        ! SEPARATE TITLE FOR MALES
DA NI=9 NO=375 MA=CM
LA
M1 M2 M3 M4 M5 M6 M7 M8 M9
KM
1.000
0.689  1.000
```

(cont.)

TABLE 7.10. *(cont.)*

```
0.204  0.218  1.000
0.335  0.284  0.315  1.000
0.274  0.320  0.153  0.265  1.000
0.333  0.333  0.221  0.364  0.268  1.000
0.258  0.211  0.114  0.139  0.185  0.132  1.000
0.319  0.346  0.176  0.207  0.231  0.279  0.146  1.000
0.316  0.269  0.111  0.140  0.117  0.131  0.263  0.163  1.000
SD
1.598  2.018  2.094  2.232  2.108  2.113  2.286  2.174  1.788
MO NX=9 NK=1 PH=PS LX=PS TD=PS                    ! PROGRAMMING SHORTCUTS
OU ME=ML RS MI SC AD=OFF IT=100 ND=4
```

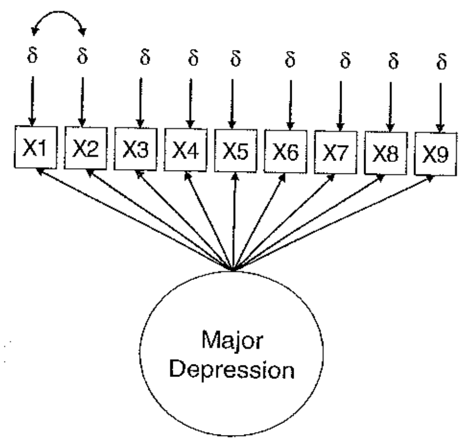
Mplus 3.11

```
TITLE:       MPLUS PROGRAM FOR EQUAL FORM OF MAJOR DEPRESSION
DATA:        FILE IS "C:\MDDALL.DAT";      ! DATA READ FROM RAW DATA FILE
VARIABLE:    NAMES ARE SUBJ SEX M1-M9;
             USEVAR ARE M1-M9;
             GROUPING IS SEX (0=FEMALE 1=MALE); ! SPECIFY GROUPING FACTOR
ANALYSIS:    ESTIMATOR=ML;
MODEL:       DEPRESS BY M1-M9;
             M1 WITH M2;
MODEL MALE:  DEPRESS BY M2-M9;               ! FREELY ESTIMATE PARAMETERS
             M1 WITH M2;                     ! IN MEN AS WELL AS IN WOMEN
OUTPUT:      SAMPSTAT MODINDICES(10.00) STAND RESIDUAL;
```

EQS 5.7b

```
/TITLE
 EQS SYNTAX FOR EQUAL FORM OF MAJOR DEPRESSION (FEMALES)
/SPECIFICATIONS
 CASES=375; VAR=9; ME=ML; MA=COR; ANALYSIS=COV; GROUPS=2;
                                                      ! NOTE: GROUPS=2
/MATRIX
1.000
0.616  1.000
0.315  0.313  1.000
0.349  0.332  0.261  1.000
0.314  0.250  0.270  0.327  1.000
0.418  0.416  0.298  0.328  0.317  1.000
0.322  0.313  0.096  0.117  0.130  0.140  1.000
0.409  0.415  0.189  0.314  0.303  0.281  0.233  1.000
0.318  0.222  0.051  0.115  0.140  0.150  0.217  0.222  1.000
/STANDARD DEVIATIONS
1.717 2.015 2.096 2.212 2.132 2.005 2.062 2.156 1.791
/LABELS
 V1=depmood; V2=anhedon; V3=weight; V4=sleep; V5=motor; V6=fatigue;
 V7=guilt; V8=concent; V9=suicide;
 F1 = DEPRESS;
/EQUATIONS
 V1 = F1+E1;
 V2 = *F1+E2;
```

(cont.)



Females: Sample Correlations, Means (M), and Standard Deviations (SD); $N = 375$

	MDD1	MDD2	MDD3	MDD4	MDD5	MDD6	MDD7	MDD8	MDD9
MDD1	1.000								
MDD2	0.616	1.000							
MDD3	0.315	0.313	1.000						
MDD4	0.349	0.332	0.261	1.000					
MDD5	0.314	0.250	0.270	0.327	1.000				
MDD6	0.418	0.416	0.298	0.328	0.317	1.000			
MDD7	0.322	0.313	0.096	0.117	0.130	0.140	1.000		
MDD8	0.409	0.415	0.189	0.314	0.303	0.281	0.233	1.000	
MDD9	0.318	0.222	0.051	0.115	0.140	0.150	0.217	0.222	1.000
M:	4.184	3.725	1.952	3.589	2.256	3.955	3.869	3.595	1.205
SD:	1.717	2.015	2.096	2.212	2.132	2.005	2.062	2.156	1.791

Males: Sample Correlations, Means (M), and Standard Deviations (SD); $N = 375$

	MDD1	MDD2	MDD3	MDD4	MDD5	MDD6	MDD7	MDD8	MDD9
MDD1	1.000								
MDD2	0.689	1.000							
MDD3	0.204	0.218	1.000						
MDD4	0.335	0.284	0.315	1.000					
MDD5	0.274	0.320	0.153	0.265	1.000				
MDD6	0.333	0.333	0.221	0.364	0.268	1.000			
MDD7	0.258	0.211	0.114	0.139	0.185	0.132	1.000		
MDD8	0.319	0.346	0.176	0.207	0.231	0.279	0.146	1.000	
MDD9	0.316	0.269	0.111	0.140	0.117	0.131	0.263	0.163	1.000
M:	4.171	3.685	1.739	3.357	2.235	3.661	3.421	3.517	1.259
SD:	1.598	2.018	2.094	2.232	2.108	2.113	2.286	2.174	1.788

FIGURE 7.3. Measurement model of DSM-IV Major Depression. MDD1, depressed mood; MDD2, loss of interest in usual activities; MDD3, weight/appetite change; MDD4, sleep disturbance; MDD5, psychomotor agitation/retardation; MDD6, fatigue/loss of energy; MDD7, feelings of worthlessness/guilt; MDD8, concentration difficulties; MDD9, thoughts of death/suicidality.

TABLE 7.9. Tests of Measurement Invariance and Population Heterogeneity of DSM-IV Major Depressive Disorder in Men and Women

	χ^2	df	χ^2_{diff}	Δdf	RMSEA (90% CI)	CFit[a]	SRMR	CFI	TLI
Single Group Solutions									
Men (n = 375)	45.96**	26			.045 (.022–.066)	.62	.041	.97	0.96
Women (n = 375)	52.95**	26			.053 (.032–.073)	.39	.044	.96	0.95
Measurement Invariance									
Equal form	98.91***	52			.049 (.034–.064)		.043	.96	0.95
Equal factor loadings	102.84***	60	3.93	8	.043 (.029–.056)		.045	.96	0.96
Equal indicator intercepts	115.31***	68	12.47	8	.043 (.029–.056)		.045	.96	0.96
Equal indicator error variances	125.02***	77	9.71	9	.041 (.027–.053)		.048	.96	0.97
Population Heterogeneity									
Equal factor variance	125.81***	78	0.79	1	.040 (.027–.053)		.050	.96	0.97
Equal latent mean	127.73***	79	1.92	1	.041 (.027–.053)		.052	.96	0.97

Note. $N = 750$. χ^2_{diff}, nested χ^2 difference; RMSEA, root mean square error of approximation; 90% CI, 90% confidence interval for RMSEA; CFit, test of close fit (probability RMSEA ≤ .05); SRMR, standardized root mean square residual; CFI, comparative fit index; TLI, Tucker–Lewis Index.
[a]Mplus 3.11 does not provide the test of close fit in multiple-groups solutions.
** $p < .01$, *** $p < .001$.

the equality of indicator intercepts; (5) test the equality of indicator residual variances (optional); and, if substantively meaningful; (6) test the equality of factor variances; (7) test the equality of factor covariances (if applicable, i.e., > 1 latent factor); and (8) test the equality of latent means. Steps 1–5 are tests of measurement invariance; Steps 6–8 are tests of population heterogeneity.

Finally, some debate continues as to whether multiple-groups CFA should be prefaced by an overall test of covariance matrices across groups. Introduced by Jöreskog (1971b), the rationale of this procedure is that if group differences exist in the parameters of the CFA model, then some values within the covariance matrix should also differ across groups. If the overall test of the equality of covariances fails to reject the null hypothesis, no further analyses are conducted and it is concluded that the groups are invariant. Rejection of the null hypothesis (e.g., $\Sigma_1 \neq \Sigma_2$) is interpreted as justification for conducting multiple-groups CFA to identify the source(s) of noninvariance (for an illustration of how to employ this test, see Vandenberg & Lance, 2000). Although some researchers have supported its continued use (e.g., Vandenberg & Lance, 2000), many methodologists have questioned the rationale and utility of the omnibus test of equal covariance matrices (e.g., Byrne, 1998; Byrne et al., 1989; Jaccard & Wan, 1996). For instance, Byrne (1998) notes that this test often produces contradictory findings with respect to equivalencies across groups; that is, occasions where the omnibus test indicates $\Sigma_1 = \Sigma_2$ but subsequent hypothesis tests of the invariance of specific CFA measurement or structural parameters must be rejected, and vice versa. Jaccard and Wan (1996) add that if the researcher has a specific hypothesis regarding group differences on selected parameters, it is better to proceed directly to the multiple-groups CFA framework because this more focused test will have greater statistical power than the omnibus comparison of covariance matrices. Accordingly, these methodologists have concluded that the omnibus test of equal covariance matrices provides little guidance for testing the equivalence of CFA parameters and thus should not be regarded as a prerequisite to multiple-groups CFA.

The multiple-groups CFA methodology is now illustrated using an actual data set of 750 adult outpatients (375 men, 375 women) with current mood disorders. In this example, the researcher is interested in examining the generalizability of the DSM-IV (American Psychiatric Association, 1994) criteria for the diagnostic category of major depressive disorder (MDD) between sexes. The analysis was motivated by questions that had arisen in the field regarding the possibility of salient sex differences in the expression of mood disorders (e.g., somatic symptoms such as appetite/weight change may be more strongly related to depression in women). Patients were rated by experienced clinicians on the severity of the nine symptoms that comprise the diagnostic criteria of MDD on 0–8 scales (0 = none, 8 = very severely disturbing/disabling; see Figure 7.3 for description of the 9 symptoms). Prior to the CFAs, the data were screened to ensure their suitability for the ML estimator (i.e., normality, absence of multivariate outliers). In accord with its DSM-IV conceptualization, a unidimensional model of MDD was posited (see Figure 7.3). For substantive reasons (cf. DSM-IV), a correlated residual was specified between the first two diagnostic criteria (i.e., depressed mood, loss of interest in usual activities). The first criterion (M1, depressed mood) was used as a marker indicator to define the metric of the latent variable. Accordingly, the covariance structure aspect of the model was overidentified in both groups with $df = 26$ (45 variances/covariances, 19 freely estimated parameters).

Prior to conducting the multiple-groups CFAs, it is important to ensure that the posited one-factor model is acceptable in both groups. If markedly disparate measurement models are obtained between groups, this outcome would contraindicate further invariance evaluation. As shown in Table 7.9, in both men and women, overall fit statistics for the one-factor solution were consistent with good model fit. In both groups, all freely estimated factor loadings were statistically significant (all $ps < .001$) and salient (completely standardized factor loadings ranged from .34 to .73). No remarkable points of strain were noted in either solution, as reflected by small modification indices, expected parameter change values, and standardized residuals.

Next, the simultaneous analysis of equal form was conducted. Table 7.10 provides the LISREL, Mplus, Amos, and EQS syntax for the equal form analysis. In LISREL, Amos, and EQS, the programming essentially entails "stacking" the CFA analysis of one group on top of the other. In LISREL and EQS, the programs are alerted to the multiple groups analysis by an additional command on the second line of syntax (NG = 2 on the DATA line of LISREL; GROUPS = 2 on the /SPECIFICATIONS line of EQS). What follows in the EQS programming is stacking the two CFA analyses for men and women; that is, the programming is identical to the syntax that would be written for conducting the CFA in a single group. The LISREL programming also follows this logic, but some programming shortcuts are implemented. Note that on the MODEL (MO) line for men, the keyword value "PS" is used in the specification of the lambda-X (LX), phi (PH), and theta-delta (TD) matrices. In LISREL, "PS" is used to signify

Conversely, MIMIC models entail the analysis of a single covariance matrix that, in addition to the indicators, includes the dummy code(s) that convey group membership. MIMIC, an acronym for "multiple indicators, multiple causes," has also been referred to as *CFA with covariates*. In this approach, both the latent factors and indicators are regressed onto dummy code(s) that denote group membership. A significant direct effect of the dummy code (covariate) on the latent factor indicates population heterogeneity (group differences on latent means) and a significant direct effect of the dummy code on an indicator is evidence of measurement noninvariance (group differences on the indicator's intercept, i.e., *differential item functioning*). Because a single input matrix is used, the advantages of MIMIC models over multiple-groups CFA include their greater parsimony (MIMIC entails fewer freely estimated parameters), their relatively greater ease of implementation when several groups are involved (i.e., depending on the complexity of the measurement model, multiple-groups CFA may be cumbersome when the number of groups exceeds two), and their less restrictive sample size requirements (i.e., multiple-groups CFA requires a sufficiently large sample size for each group). However, the key limitation of MIMIC models relative to multiple-groups CFA is their ability to examine just two potential sources of invariance (indicator intercepts, factor means).

Multiple-Groups CFA

Before considering how multiple-groups CFA is implemented, it is important to be aware of the variability in terminologies and procedures within the methodology literature. As noted above, an advantage of multiple-groups CFA is that all potential aspects of invariance across groups can be examined. Different terminologies exist in the literature for these various tests of invariance (cf. Horn & McArdle, 1992; Meredith, 1993). For example, the test of equal factor structures ("equal form," meaning that the number of factors and pattern of indicator–factor loadings is identical across groups) has been referred to as "configural invariance." Equality of factor loadings has been referred to as "metric invariance" or "weak factorial invariance." The equality of indicator intercepts has been alternatively termed "scalar invariance" or "strong factorial invariance." Finally, evaluation of the equality of indicator residuals has also been referred to as a test of "strict factorial invariance" (Meredith, 1993). Some confusion surrounds the use of these varying terminologies (cf. Vandenberg & Lance, 2000). For this reason, a more descriptive and pedagogically useful terminology is encouraged and used in this book (e.g., equal form, equal factor loadings, equal intercepts).

Moreover, there are some discrepancies in the order in which the model restrictions within multiple-groups CFA are evaluated. Most commonly, the stepwise procedures previously illustrated in the tests of longitudinal invariance are employed, whereby the analysis begins with the least restricted solution (equal form) and subsequent models are evaluated (using nested χ^2 methods) that entail increasingly restrictive constraints; that is, equal factor loadings → equal intercepts → equal residual variances, and so on. However, some methodologists (e.g., Horn & McArdle, 1992) have proffered a "step-down" strategy whereby the starting model contains all the pertinent invariance restrictions, and subsequent models are then evaluated that sequentially relax these constraints. The former approach is recommended for several reasons. Especially in the context of a complex CFA solution (i.e., multiple factors and indicators; > 2 groups), it may be difficult to determine the (multiple) sources of noninvariance when a model held to full invariance is poor fitting. Any aspects of ill fit that are encountered are more easily identified and adjusted for in a model-building approach, where new restrictions are placed on the solution at each step. In addition, tests of some aspects of invariance rest on the assumption that other aspects of invariance hold. Group comparisons of latent means are meaningful only if the factor loadings and indicator intercepts have been found to be invariant (see the "Longitudinal Measurement Invariance" section). Group comparisons of factor variances and covariances are meaningfully only when the factor loadings are invariant. The viability of the fully constrained model rests on the results of the less restricted solutions. Thus, it is more prudent for model evaluation to work upward from the least restricted solution (equal form) to determine if further tests of measurement invariance and population heterogeneity are warranted. Of relevance here is the issue of *partial invariance*. As discussed later in this chapter, it may be possible for the evaluation of group equivalence to proceed even in instances where some noninvariant parameters have been encountered (cf. Byrne et al., 1989). For instance, it may be possible to compare groups on latent means if some (but not all) of the factor loadings and intercepts are invariant. Again, a "step-up" approach to invariance evaluation would foster this endeavor.

For the aforementioned reasons, the recommended sequence of multiple-groups CFA invariance evaluation is as follows: (1) Test the CFA model separately in each group; (2) conduct the simultaneous test of equal form (identical factor structure); (3) test the equality of factor loadings; (4) test

Heterogeneity of variance is a common outcome in repeated measures designs, such as the current example. Thus, the test of equal residual variances usually fails in actual data sets because of the temporal fanspread of indicator variances. In the present context, this could be reflective of individual differences in response to the intervention to improve job satisfaction. That is, at Time 1 the variances were more homogeneous because individuals were more similar with regard to their level of job satisfaction. By Time 2, individual differences were more pronounced because some participants responded favorably to the intervention whereas others did not. This is reflected in the input matrix, where it can be seen that the *SD*s increase in magnitude from Time 1 to Time 2 (in addition to the *M*s, which reflect overall improvement in satisfaction). Accordingly, methodologists concede that the test of equal indicator residual variances is highly stringent and will rarely hold in realistic data sets (e.g., Chan, 1998). Fortunately, this condition is not as important to the evaluation of measurement invariance as the prior tests (equal form, factor loadings, and intercepts).

CFA IN MULTIPLE GROUPS

Overview of Multiple-Groups Solutions

The themes introduced in this chapter are now extended to the simultaneous analysis of more than one group. As noted in previous chapters, one of the major advantages of CFA over EFA is its capability to examine the equivalence of all measurement and structural parameters of the factor model across multiple groups. The measurement model pertains to the measurement characteristics of the indicators (observed measures) and thus consists of the factor loadings (lambda), intercepts (tau), and residual variances (theta). Hence, the evaluation of across-group equivalence of these parameters reflects tests of *measurement invariance*. The structural parameters of the CFA model involve evaluation of the latent variables themselves, and thus consist of the factor variances (phi), covariances (phi), and latent means (kappa). If latent-Y terminology is used, the corresponding matrices would be psi and alpha (cf. Figure 3.4, Chapter 3). These parameters describe characteristics of the population from which the sample was drawn. Thus, the examination of the group concordance of structural parameters can be considered tests of *population heterogeneity*; that is, do the dispersion, interrelationships, and levels of the latent factors vary across groups?

CFA with multiple groups has many potential practical applications. For instance, the issues addressed in measurement invariance evaluation are key to the psychometric development of psychological tests; for example, do the items of a questionnaire measure the same constructs (same factor structure) and evidence equivalent relationships to these constructs (equal factor loadings) in all subgroups of the population for whom the measure will be used? Or are there sex, ethnic/racial, age, or other subgroup differences that preclude responding to the questionnaire in comparable ways? Does the questionnaire contain items that are biased against a particular subgroup; that is, yield substantially higher or lower observed scores in a group at equivalent levels of the latent or "true" score? The evaluation of measurement invariance is also important to determining the generalizability of psychological constructs across groups; for example, does the construct underlying the formal definition of a given psychiatric diagnosis operate equivalently across cultures, sexes, age groups, and so forth? Tests of structural parameters reveal potential group differences adjusting for measurement error and an error theory. For example, tests of equality of factor covariances can be construed as the CFA counterpart to inferential evaluation of the differential magnitude of independent correlations; that is, are two constructs more strongly correlated in one group than another? Tests of the equality of latent means are analogous to the comparison of observed group means via *t*-test or ANOVA. However, the major strength of the CFA-based approach is that such comparisons are made in the context of a latent variable measurement model, which hence adjusts for measurement errors, correlated residuals, and so forth.

Two methods can be used to evaluate CFA solutions in multiple groups: (1) *multiple-groups CFA*; and (2) *MIMIC modeling*. Multiple-groups CFA entails the simultaneous analysis of CFA in more than one group. For instance, if the analysis involves two groups (e.g., males and females), two *separate* input matrices are analyzed and constraints can be placed on like parameters (e.g., factor loadings) in both groups to examine the equivalence of the measurement (measurement invariance) and structural solution (population heterogeneity). Although somewhat underutilized in applied research (cf. Vandenberg & Lance, 2000), multiple-groups CFA can entail the analysis of mean structures to evaluate the equality of indicator intercepts (measurement invariance) and latent means (population heterogeneity). A key advantage of multiple-groups CFA is that all aspects of measurement invariance and population heterogeneity can be examined (i.e., factor loadings, intercepts, residual variances, factor variances, factor covariances, latent means).

TABLE 7.8. Computer Syntax (LISREL, Mplus) for Testing a Fully Invariant Longitudinal Measurement Model of Job Satisfaction (Equal Form, Equal Factor Loadings, Equal Intercepts, Equal Residual Variances)

Mplus 3.11

```
TITLE:     MPLUS PROGRAM FOR TIME1-TIME2 MSMT MODEL OF JOB SATISFACTION
DATA:      FILE IS "C:\INPUT.dat";
           TYPE IS MEANS STD CORR; ! INDICATOR MEANS ALSO INPUTTED
           NOBS ARE 250;
VARIABLE:  NAMES ARE A1 B1 C1 D1 A2 B2 C2 D2;
ANALYSIS:  ESTIMATOR=ML;
           TYPE=MEANSTRUCTURE; ! ANALYSIS OF MEAN STRUCTURE
MODEL:     SATIS1 BY A1 B1 (1)
           C1 (2)
           D1 (3);
           SATIS2 BY A2 B2 (1)
           C2 (2)
           D2 (3);
           A1 WITH A2; B1 WITH B2; C1 WITH C2; D1 WITH D2;
           [A1@0]; [A2@0]; ! FIXES THE A INDICATOR INTERCEPTS TO ZERO
           [SATIS1*]; [SATIS2*]; ! FREELY ESTIMATES FACTOR MEANS
           [B1 B2] (4); [C1 C2] (5); [D1 D2] (6);
           A1 A2 (7); B1 B2 (8); C1 C2 (9); D1 D2 (10);
OUTPUT:    SAMPSTAT MODINDICES(4.00) STAND RESIDUAL;
```

LISREL 8.72

```
TITLE LISREL PROGRAM FOR TIME1-TIME2 MSMT MODEL OF JOB SATISFACTION
DA NI=8 NO=250 MA=CM
LA
A1 B1 C1 D1 A2 B2 C2 D2
KM
1.000
0.736 1.000
0.731 0.648 1.000
0.771 0.694 0.700 1.000
0.685 0.512 0.496 0.508 1.000
0.481 0.638 0.431 0.449 0.726 1.000
0.485 0.442 0.635 0.456 0.743 0.672 1.000
0.508 0.469 0.453 0.627 0.759 0.689 0.695 1.000
ME                                            ! MEANS INCLUDED AS INPUT
1.50 1.32 1.45 1.41 6.60 6.42 6.56 6.31
SD
1.94 2.03 2.05 1.99 2.61 2.66 2.59 2.55
MO NX=8 NK=2 PH=SY,FR LX=FU,FR TD=SY,FR TX=FR KA=FR     ! TAU-X AND KAPPA
LK
SATIS1 SATIS2
PA LX
0 0
1 0
```

(cont.)

TABLE 7.8. *(cont.)*

```
1 0
1 0
0 0
0 1
0 1
0 1
VA 1.0 LX(1,1) LX(5,2)             ! SET THE METRIC OF THE LATENT VARIABLES
PA TD
1
0 1
0 0 1
0 0 0 1
1 0 0 0 1                          ! OFF-DIAGONAL 1s ARE CORRELATED ERRORS
0 1 0 0 0 1
0 0 1 0 0 0 1
0 0 0 1 0 0 0 1
PA PH
1
1 1
FI TX(1) TX(5)
VA 0.0 TX(1) TX(5)                 ! FIX INDICATOR A INTERCEPTS TO ZERO
EQ LX(2,1) LX(6,2)                 ! FACTOR LOADING EQUALITY CONSTRAINT
EQ LX(3,1) LX(7,2)                 ! FACTOR LOADING EQUALITY CONSTRAINT
EQ LX(4,1) LX(8,2)                 ! FACTOR LOADING EQUALITY CONSTRAINT
EQ TX(2) TX(6)                     ! INTERCEPT EQUALITY CONSTRAINT
EQ TX(3) TX(7)                     ! INTERCEPT EQUALITY CONSTRAINT
EQ TX(4) TX(8)                     ! INTERCEPT EQUALITY CONSTRAINT
EQ TD(1,1) TD(5,5)                 ! ERROR VARIANCE EQUALITY CONSTRAINT
EQ TD(2,2) TD(6,6)                 ! ERROR VARIANCE EQUALITY CONSTRAINT
EQ TD(3,3) TD(7,7)                 ! ERROR VARIANCE EQUALITY CONSTRAINT
EQ TD(4,4) TD(8,8)                 ! ERROR VARIANCE EQUALITY CONSTRAINT
OU ME=ML RS MI SC AD=OFF IT=100 ND=4
```

Note. In Mplus, there can be only one number in parentheses on each line. Recall that parameters that are followed by the same number in parentheses are constrained to be equal. Thus, the syntax (1), (2), and (3) hold the factor loadings of B1 and B2, C1 and C2, and D1 and D3 to equality (although A1 and A2 are on the same line as B1 and B2, these parameters are not constrained to equality because they have been fixed to 1.0 by Mplus default). The Mplus language uses brackets, [ ], to represent indicator intercepts and factor means. Thus, the command [B1 B2] (4) holds the intercepts of indicators B1 and B2 to equality. The last line on the MODEL: section of the syntax, for example, A1 A2 (7), instructs the analysis to constrain the indicator error variances to equality.

TABLE 7.7. Longitudinal Invariance of a Measurement Model of Job Satisfaction (N = 250)

	χ^2	df	χ^2_{diff}	Δdf	RMSEA (90% CI)	CFit	SRMR	CFI	TLI
Equal form	2.09	15			.000 (.000–.000)	1.00	.010	1.00	1.01
Equal factor loadings	3.88	18	1.79	3	.000 (.000–.000)	1.00	.014	1.00	1.01
Equal indicator intercepts	7.25	21	3.37	3	.000 (.000–.000)	1.00	.026	1.00	1.01
Equal indicator error variances	90.73***	25	83.48***	4	.103 (.080–.126)	0.00	.037	.96	0.96

Note. χ^2_{diff}, nested χ^2 difference; RMSEA, root mean square error of approximation; 90% CI, 90% confidence interval for RMSEA; CFit, test of close fit (probability RMSEA ≤ .05); SRMR, standardized root mean square residual; CFI, comparative fit index; TLI, Tucker–Lewis Index.
***p < .001.

intercepts were invariant between the two testing occasions. Recall that in multiple regression, an intercept can be regarded as the predicted value of Y when X is at zero. Measurement intercepts in CFA can be interpreted in an analogous manner; that is, τ_x = the predicted value of indicator X when κ = 0. Thus, if the intercept of an indicator is found to be temporally *noninvariant*, the predicted score of the indicator will vary across time at a constant level of the latent construct (i.e., $\kappa_1 = \kappa_2$). Stated another way, even when the "true score" (latent factor) remains unchanged, the observed scores of the indicators will vary over time. Thus, it is erroneous to interpret changes in observed scores as true change (alpha change), because the observed change is due in some part to temporal variation in the measurement properties of the indicator. For example, although factor loading equivalence would suggest the indicator possesses a temporally stable relationship to the underlying construct (i.e., a unit increase in the latent construct is associated with comparable changes in the indicator at all assessment points), noninvariant indicator intercepts would suggest inequality of the indicator's location parameters over time (a spurious shift from using one portion of the indicator's response scale at Time 1 to another portion of the response scale at Time 2, as might occur in various forms of rater drift such as leniency bias; cf. Vandenberg & Lance, 2000). However, in the present illustration both the factor loadings and indicator intercepts were found to be invariant, suggesting the analysis of mean change over time can be attributed to true change in the construct (cf. Eq. 7.7).

The final analysis tests for the equality of the indicator's error variances. Table 7.8 provides Mplus and LISREL syntax for this analysis. This restriction results in a significant decrease in model fit, χ^2_{diff} (4) = 83.48, p < .001 (critical χ^2 value at df = 4, α = .05, is 9.49). Fit diagnostics suggest that each indicator's error variance is temporally noninvariant:

```
MODEL MODIFICATION INDICES
Variances/Residual Variances

                M.I.    E.P.C.   Std E.P.C.  StdYX E.P.C.
A1            35.185    -0.421       -0.421        -0.108
B1            24.357    -0.579       -0.579        -0.125
C1            15.381    -0.421       -0.421        -0.093
D1            23.308    -0.459       -0.459        -0.112
A2            35.187     0.474        0.474         0.076
B2            24.358     0.602        0.602         0.087
C2            15.381     0.442        0.442         0.065
D2            23.307     0.483        0.483         0.078
```

TABLE 7.6. Mplus Results of the Equal Form Longitudinal Model of Job Satisfaction

MODEL RESULTS

	Estimates	S.E.	Est./S.E.	Std	StdYX
SATIS1 BY					
A1	1.000	0.000	0.000	1.728	0.899
B1	0.951	0.050	19.212	1.643	0.811
C1	0.970	0.049	19.616	1.676	0.814
D1	0.990	0.046	21.394	1.710	0.859
SATIS2 BY					
A2	1.000	0.000	0.000	2.327	0.896
B2	0.916	0.048	18.973	2.131	0.808
C2	0.922	0.046	20.085	2.144	0.828
D2	0.934	0.045	20.770	2.173	0.849
SATIS2 WITH					
SATIS1	2.680	0.353	7.586	0.667	0.667
A1 WITH					
A2	0.701	0.119	5.915	0.701	0.141
B1 WITH					
B2	1.043	0.163	6.407	1.043	0.195
C1 WITH					
C2	1.042	0.158	6.576	1.042	0.195
D1 WITH					
D2	0.770	0.134	5.735	0.770	0.151
Means					
SATIS1	1.500	0.122	12.343	0.868	0.868
SATIS2	6.600	0.164	40.183	2.836	2.836
Intercepts					
A1	0.000	0.000	0.000	0.000	0.000
B1	-0.107	0.117	-0.910	-0.107	-0.053
C1	-0.005	0.118	-0.042	-0.005	-0.002
D1	-0.075	0.108	-0.692	-0.075	-0.038
A2	0.000	0.000	0.000	0.000	0.000
B2	0.375	0.340	1.102	0.375	0.142
C2	0.477	0.324	1.475	0.477	0.184
D2	0.146	0.316	0.462	0.146	0.057

(cont.)

TABLE 7.6. *(cont.)*

Variances					
SATIS1	2.985	0.321	9.297	1.000	1.000
SATIS2	5.414	0.586	9.244	1.000	1.000
Residual Variances					
A1	0.707	0.105	6.722	0.707	0.192
B1	1.409	0.153	9.178	1.409	0.343
C1	1.434	0.157	9.137	1.434	0.338
D1	1.038	0.127	8.165	1.038	0.262
A2	1.330	0.195	6.815	1.330	0.197
B2	2.418	0.263	9.183	2.418	0.347
C2	2.113	0.239	8.856	2.113	0.315
D2	1.836	0.219	8.390	1.836	0.280

for the equality of factor loadings over the two assessment points. Unlike the evaluation of tau equivalence, this analysis does not constrain the indicators that load on the same factor to equality (e.g., $\lambda_{21} = \lambda_{31} = \lambda_{41}$). Rather, the equality constraint pertains to the factor loadings of indicators administered repeatedly across testing occasions (e.g., $\lambda_{21} = \lambda_{62}$). As in the analysis of tau equivalence, the nested χ^2 test can be employed to determine whether these constraints significantly degrade model fit. As shown in Table 7.7, the model χ^2 of the equal loadings solution is 3.88 ($df = 18$, $p = 1.0$), resulting in a nonsignificant χ^2 difference test, χ^2_{diff} (3) = 1.79, *ns*; [critical value of $\chi^2(3) = 7.81$, $\alpha = .05$]. The *df* of this nested model comparison is equal to 3 (and the equal factor loadings model $df = 18$) because only three pairs of factor loadings (B1 = B2; C1 = C2; D1 = D2) are involved in this equality constraint; that is, does not apply to the loadings of A1 and A2 because these parameters were previously fixed to 1.0 to set the scale of the latent factors.

On the basis of the results of the equal factor loading analysis, it can be concluded that the indicators evidence equivalent relationships to the latent construct of Job Satisfaction over time. Keeping the equality constraints of the factor loadings in place, the next model placed additional equality constraints on the indicators' intercepts, except for indicators A1 and A2, whose intercepts were previously fixed to zero for the purposes of model identification. These restrictions also did not lead to a significant reduction in model fit, χ^2_{diff} (3) = 3.37, *ns*, suggesting that the indicator's

TABLE 7.6. Mplus Results of the Equal Form Longitudinal Model of Job Satisfaction

MODEL RESULTS

	Estimates	S.E.	Est./S.E.	Std	StdYX
SATIS1 BY					
A1	1.000	0.000	0.000	1.728	0.899
B1	0.951	0.050	19.212	1.643	0.811
C1	0.970	0.049	19.616	1.676	0.814
D1	0.990	0.046	21.394	1.710	0.859
SATIS2 BY					
A2	1.000	0.000	0.000	2.327	0.896
B2	0.916	0.048	18.973	2.131	0.808
C2	0.922	0.046	20.085	2.144	0.828
D2	0.934	0.045	20.770	2.173	0.849
SATIS2 WITH					
SATIS1	2.680	0.353	7.586	0.667	0.667
A1 WITH					
A2	0.701	0.119	5.915	0.701	0.141
B1 WITH					
B2	1.043	0.163	6.407	1.043	0.195
C1 WITH					
C2	1.042	0.158	6.576	1.042	0.195
D1 WITH					
D2	0.770	0.134	5.735	0.770	0.151
Means					
SATIS1	1.500	0.122	12.343	0.868	0.868
SATIS2	6.600	0.164	40.183	2.836	2.836
Intercepts					
A1	0.000	0.000	0.000	0.000	0.000
B1	-0.107	0.117	-0.910	-0.107	-0.053
C1	-0.005	0.118	-0.042	-0.005	-0.002
D1	-0.075	0.108	-0.692	-0.075	-0.038
A2	0.000	0.000	0.000	0.000	0.000
B2	0.375	0.340	1.102	0.375	0.142
C2	0.477	0.324	1.475	0.477	0.184
D2	0.146	0.316	0.462	0.146	0.057

(cont.)

TABLE 7.6. *(cont.)*

Variances					
SATIS1	2.985	0.321	9.297	1.000	1.0C
SATIS2	5.414	0.586	9.244	1.000	1.0C
Residual Variances					
A1	0.707	0.105	6.722	0.707	0.19
B1	1.409	0.153	9.178	1.409	0.34
C1	1.434	0.157	9.137	1.434	0.33
D1	1.038	0.127	8.165	1.038	0.26
A2	1.330	0.195	6.815	1.330	0.19
B2	2.418	0.263	9.183	2.418	0.34
C2	2.113	0.239	8.856	2.113	0.31
D2	1.836	0.219	8.390	1.836	0.28

for the equality of factor loadings over the two assessment points. Unlik the evaluation of tau equivalence, this analysis does not constrain the indi cators that load on the same factor to equality (e.g., $\lambda_{21} = \lambda_{31} = \lambda_{41}$). Rathe the equality constraint pertains to the factor loadings of indicators admin istered repeatedly across testing occasions (e.g., $\lambda_{21} = \lambda_{62}$). As in the analy sis of tau equivalence, the nested χ^2 test can be employed to determin whether these constraints significantly degrade model fit. As shown i Table 7.7, the model χ^2 of the equal loadings solution is 3.88 (df = 18, p 1.0), resulting in a nonsignificant χ^2 difference test, χ^2_{diff} (3) = 1.79, *ns* [critical value of $\chi^2(3) = 7.81$, $\alpha = .05$]. The df of this nested model com parison is equal to 3 (and the equal factor loadings model df = 18) becaus only three pairs of factor loadings (B1 = B2; C1 = C2; D1 = D2) ar involved in this equality constraint; that is, does not apply to the loading of A1 and A2 because these parameters were previously fixed to 1.0 to se the scale of the latent factors.

On the basis of the results of the equal factor loading analysis, it ca be concluded that the indicators evidence equivalent relationships to th latent construct of Job Satisfaction over time. Keeping the equality con straints of the factor loadings in place, the next model placed additiona equality constraints on the indicators' intercepts, except for indicators A and A2, whose intercepts were previously fixed to zero for the purposes o model identification. These restrictions also did not lead to a significan reduction in model fit, χ^2_{diff} (3) = 3.37, *ns*, suggesting that the indicator tors was defined by using the same observed measure at both testing occasions (A1 and A2) as the marker indicator.

The estimation of indicator intercepts and factor means requires that the observed means of the indicators be included as input data. Thus, the number of elements of the input matrix increases by the number of indicator means that are included in the analysis; that is,

$$b = [p(p + 1) / 2] + p \tag{7.8}$$

where b is the number of elements of the input matrix and p is the number of indicators. Thus, in the current example (p = 8 indicators), there are 36 variances and covariances $[p(p + 1) / 2]$ and 8 means (p), totaling to 44 elements of the input matrix; Eq. 7.8 can be alternatively written as $b = [p(p + 3)] / 2$. Although this exceeds the number of freely estimated parameters of the equal form solution, the mean structure portion of the solution is underidentified in the absence of other model restrictions. In this example, there are 8 observed means (knowns) but 10 unknown parameters in the CFA mean structure (8 indicator intercepts, 2 latent factor means). Hence, this aspect of the model is underidentified because the number of unknowns exceeds the number of knowns (cf. Chapter 3). Moreover, just as latent variables must be provided a metric (e.g., by fixing the loading of an indicator to 1.0), they must also be assigned an origin (i.e., a mean). In covariance structure analysis, the latent variable means are assumed to be zero. In mean structure analysis, where the latent variables may take on mean values other than zero, origins must be assigned. In a single-sample analysis, the mean structure aspect of the measurement model may be identified in one of two ways: (1) fixing the latent mean to zero; or (2) assigning the factor mean to take on the mean as one of its indicators (by fixing the intercept of one indicator to zero). In the first method, all the indicator intercepts are freely estimated but will equal the observed means of the indicators. In the second strategy, all but one intercept is freely estimated and these freed intercepts will take on values different from their sample means; the factor mean is freely estimated but will equal the observed mean of the indicator whose intercept was fixed to zero. The restrictions associated with both approaches will result in just-identification of the mean structure portion of the solution. In either approach, additional restrictions can be placed on the model to over-identify its mean structure component (e.g., equality constraints on indicator intercepts). Later in this chapter, a slightly different method is used to identify the mean structure of a multiple-groups CFA solution.

In the current example, Indicator A's intercept was fixed to zero at both testing occasions. Consequently, in the initial, less constrained solutions (e.g., the test of equal factor structure), the mean of the latent factor of Job Satisfaction was equal to the observed mean of Indicator A. In addition, these restrictions led to just-identification of the mean structure aspect of the solution; that is, 8 observed means; 8 freely estimated mean parameters = 6 intercepts + 2 latent factor means. Thus, although the input matrix is expanded to include indicator means, the overall model df is the same as in covariance structure analysis (df = 15) because the mean structure component does not provide additional degrees of freedom to the model ($df = 8 - 8 = 0$).

Selected results of the equal form (factor structure) solution are presented in Table 7.6. The overall fit of this solution is presented in Table 7.7. All results are in accord with the conclusion that a unidimensional measurement model of Job Satisfaction is viable at both testing occasions. Each of the overall fit statistics is consistent with good model fit, for example, $\chi^2(15) = 2.09$, $p = 1.0$, and fit diagnostics indicate the absence of significant areas of strain in the solution (e.g., all modification indices < 4.0). At both assessments, the indicators are found to be significantly ($ps < .001$) and strongly related to the latent construct of Job Satisfaction (range of completely standardized factor loadings = .81 to .90). In addition, the four error covariances are statistically significant ($ps < .001$); completely standardized values range in magnitude from .14 to .195. The test–retest covariance of the latent construct of Job Satisfaction is statistically significant ($\phi_{21} = 2.68$, $p < .001$). As noted above, the intercept of Indicator A was fixed to zero at both assessments in order to identify the mean structure component of the solution. This is reflected by the values of 0.00 for all estimates of A1 and A2 in the Intercepts portion of the model results. As the result of these restrictions, the latent means of Job Satisfaction equal the observed means of the A1 and A2 indicators (1.5 and 6.6 for SATIS1 and SATIS2, respectively; cf. Table 7.6 and Figure 7.2). To illustrate an equation presented earlier in this chapter, the sample mean of the indicators can be reproduced by inserting the appropriate model estimates into Equation 7.7. For instance, the sample mean of B1 = 1.32 and can be reproduced using the model estimates of its intercept ($\tau_1 = -0.107$), unstandardized factor loading ($\lambda_{21} = 0.951$), and the mean of its latent factor ($\kappa = 1.5$):

$$-0.107 + 0.951(1.5) = 1.32 \tag{7.9}$$

Given evidence of equal form (cf. gamma change), additional tests of longitudinal measurement invariance may proceed. The next analysis tests from their means (i.e., their means are presumed to be zero). For example, as with coefficients in multiple regression, a factor loading can be interpreted as the amount of predicted change in an indicator (the Y variable) given a unit change in the latent factor (the X variable). These coefficients do not reflect the exact predicted score of the Y variable, but estimate how much this variable is predicted to change, given a unit change in X. However, a typical application of multiple regression involves the prediction of specific scores of the Y variable, as shown in the simple regression equation below:

$$\hat{Y} = a + bX \tag{7.1}$$

where $\hat{Y}$ is the predicted Y score; a is the intercept; b is the unstandardized regression coefficient; and X is a given score of the predictor variable. The actual (observed value) of Y can be reproduced by adding the residual (e) to the sum of $a + bX$:

$$Y = a + bX + e \tag{7.2}$$

A basic equation in regression solves for the intercept (a) on the basis of knowledge of the sample means of X and Y (M_y, M_x) and the regression coefficient (b):

$$a = M_y - bM_x \tag{7.3}$$

Through a simple algebraic manipulation, this formula can be reexpressed as

$$M_y = a + bM_x \tag{7.4}$$

A fundamental difference between these equations and the equations presented in the context of CFA thus far is its inclusion of the intercept parameter. In the multiple regression framework, the intercept represents the position where the least squares regression line crosses the Y axis; that is, the predicted value of Y when all X variables are zero. In Chapter 3 (Eq. 3.4), it was shown that the observed variance of an indicator could be calculated from the model estimates of a covariance structure CFA using the formula (Latent X notation)

$$VAR(X) = \lambda_x{}^2\phi + \delta \tag{7.5}$$

where λ_x is the unstandardized factor loading; ϕ is the factor variance; and δ is the indicator's error variance. With the exception of the intercept parameter, this equation is very similar to the previous regression equation (e.g., $\lambda_x = b$; $\delta = e$).

The equations of CFA can be expanded to include intercept (and latent mean) parameters; for example,

$$X = \tau_x + \Lambda_x\xi + \theta_\delta \tag{7.6}$$

where τ_x is the indicator intercept (τ_x = tau X; cf. Figure 3.3 in Chapter 3). Similarly, the CFA equivalent to the equation, $M_y = a + bM_x$, is

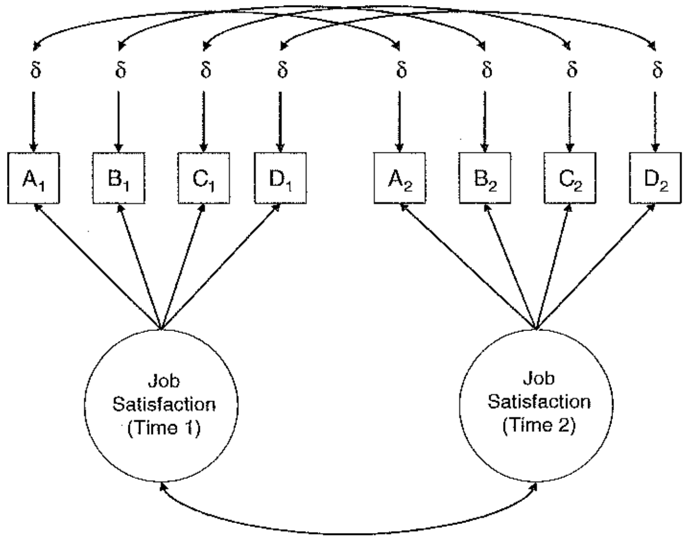
$$M_x = \tau_x + \lambda_x\kappa \tag{7.7}$$

where κ is the mean of the latent exogenous factor (κ = kappa; cf. Figure 3.3 in Chapter 3). Thus, the mean of a given indicator (M_x) can be reproduced by the CFA model's parameter estimates of the indicator's intercept (τ_x), factor loading (λ_x), and latent factor mean (κ).

Later in this chapter it will be shown that, like marker indicators and latent factor variances, indicator intercepts (means) and latent factor means are closely related. In addition, the analysis of mean structure poses new identification issues that can be addressed in different ways in the single- and multiple-group approaches.

Returning to the example of longitudinal measurement invariance, this analysis can be conducted with or without the inclusion of mean structures. For instance, if the ultimate goal is to examine models based on a covariance structure (e.g., as in autoregressive/cross-lagged panel modeling), the analysis of mean structures is less relevant. However, if the goal is to examine the trajectory of change in the level of a given construct (e.g., as in latent growth curve modeling), the measurement invariance evaluation should include the analysis of indicator means; that is, the comparison of means is meaningful only if the factor loadings and measurement intercepts are found to be invariant.

The first step of the longitudinal analysis of job satisfaction is to establish that the same factor structure is present at both testing occasions (equal form). As shown in Figure 7.2, it is predicted that a unidimensional measurement model of job satisfaction is viable at both assessment points. Correlated errors were specified in anticipation that additional covariance would exist between repeated measures owing to temporally stable indicator-specific variance (method effects). The metric of the latent fac-



Sample Correlations, Means (*M*), and Standard Deviations (*SD*); *N* = 250

	A1	B1	C1	D1	A2	B2	C2	D2
A1	1.000							
B1	0.736	1.000						
C1	0.731	0.648	1.000					
D1	0.771	0.694	0.700	1.000				
A2	0.685	0.512	0.496	0.508	1.000			
B2	0.481	0.638	0.431	0.449	0.726	1.000		
C2	0.485	0.442	0.635	0.456	0.743	0.672	1.000	
D2	0.508	0.469	0.453	0.627	0.759	0.689	0.695	1.000
M:	1.500	1.320	1.450	1.410	6.600	6.420	6.560	6.310
SD:	1.940	2.030	2.050	1.990	2.610	2.660	2.590	2.550

FIGURE 7.2. Longitudinal measurement model of job satisfaction. Indicators A through D are various measures (e.g., questionnaires, supervisor ratings) of job satisfaction. Rectangles denote within-time portions of the input correlation matrix.

examine the other aspects of measurement invariance (e.g., equality of factor loadings) in a manner similar to the procedures used for evaluating tau-equivalent and parallel indicators. In addition, note that correlated errors have been specified for each pair of repeated measures (e.g., the residual of indicator A1 is allowed to freely covary with the residual of indicator A2). Recall from earlier chapters (e.g., Chapter 2) that the uniqueness of an indicator is comprised of some combination of random measurement error and indicator-specific variance. Specification of correlated errors is based on the premise that these indicator-specific variances are temporally stable. For instance, such parameters might be posited to reflect method effects associated with repeated administrations of the same measure (e.g., in addition to the underlying dimension of job satisfaction, some of the variance in the A indicator is due to the influence of social desirability, present at each testing occasion). Depending on the research scenario, it may or may not be necessary to specify autocorrelations among indicators' error variances; for example, these effects may be less likely evident in designs employing wide assessment intervals or monomethod indicators.

It should be noted that longitudinal measurement invariance can also be evaluated using the multiple-group approach discussed later in this chapter; that is, each "group" is represented by a different wave of assessment (e.g., Group 1 = Time 1, Group 2 = Time 2). In reviewing the measurement invariance literature, Vandenberg and Lance (2000) discussed the advantages and disadvantages of assessing longitudinal measurement invariance using a single sample (all assessment waves combined in a single input matrix) versus a multiple-group approach (assessment waves represented by separate input matrices for each "group"). The one-sample approach takes into account the complete data structure; that is, the lagged relationships among indicators in addition to the within-time covariances. The entire matrix presented in Figure 7.2 is used in the analysis including the between-time correlations, such as A1 with A2, B2, C2, and D2. In the multiple-group approach, only the within-time portions of the matrix (denoted by boxes in Figure 7.2) are input to the analysis (each box represents a different "group"). Accordingly, a primary advantage of employing a one-sample approach is that correlated errors of the repeated measurements can be estimated and controlled for in the estimation of other model parameters. The main disadvantage of the single-sample method pertains to the use of an input matrix larger than that used in the multiple-group approach (cf. Figure 7.2). Because a larger matrix is used as input, the single-sample approach may be more prone to poor model fit or possibly improper solutions resulting from the greater complexity of the model (cf. Chapter 5).

The longitudinal measurement example is also used to introduce the reader to the analysis of mean structures. All examples been presented in this book thus far have entailed the analysis of covariance structures. That is, only variances and covariances have been included in the input matrix, and thus the indicators and resulting model parameters were deviations cussion continues with the next section, a couple of additional points are made about the previous analyses. First, the nested χ^2 procedure was employed by comparing the model in question to the previous, slightly less restricted solution; for example, a model in which the indicators of both Auditory Memory and Visual Memory are held to be parallel versus the model in which only the indicators of Auditory Memory are held parallel. It should be noted that any two models presented in Table 7.3 are nested and thus any less restricted solution could serve as the comparison model. For instance, the final solution (in which both the indicators of Auditory Memory and Visual Memory are parallel) could be tested against the congeneric solution. In this case, the χ^2_{diff} value would be 4.40 with $df = 8$; that is, $\chi^2 = 9.28 - 4.88 = 4.40$, $df = 16 - 8 = 8$. This comparison would produce the equivalent result that the fully parallel model did not significantly degrade the fit of the solution because the χ^2 difference of 4.40 is less than the critical value of the χ^2 distribution at $df = 8$ (i.e., $\chi^2_{\text{crit}} = 15.51$, at $\alpha = .05$). It might appear that it would be more efficient to move straight from the congeneric model to the most restrictive solution. However, the conditions of tau-equivalent and parallel indicators often do not hold in applied data sets (in particular, the condition of parallel indicators is quite restrictive). Thus, it is usually better to employ an incremental strategy that will allow one to more readily detect the sources on noninvariance if significant degradations in model fit are encountered, because the restrictions are placed on a single set of parameters at a time rather than all at once.

Second, in the case of a multifactorial, congeneric solution, the evaluation of tau-equivalent and parallel indicators of one factor does not rely on the respective findings for indicators loading on different factors. For instance, if the condition of tau equivalence for the indicators of Auditory Memory was not met (i.e., these equality constraints led to a significant increase in model χ^2), the researcher could still proceed to evaluating whether the indicators of Visual Memory were tau equivalent and parallel.

Longitudinal Measurement Invariance

Another type of invariance evaluation that can be conducted on CFA models within a single group concerns the equality of construct measurement over time. Although rarely addressed in the applied literature, longitudinal measurement invariance is a fundamental aspect of evaluating temporal change in a construct. In the absence of such evaluation, it cannot be determined whether temporal change observed in a construct is due to true change or to changes in the structure or measurement of the construct

over time. Drawing on the work of Golembiewski, Billingsley, and Yeager (1976), Chan (1998) outlined three types of change that may be encountered in repeated measurements: *alpha*, *beta*, and *gamma* change (these terms do not correspond to parameters of structural equation models). Alpha change refers to true score change in a construct given a constant conceptual domain and constant measurement. Alpha change (true score change) can only be said to occur in the context of longitudinal measurement invariance (i.e., evidence that the measurement of the construct does not change over time). Chan (1998) notes that longitudinal measurement invariance can be construed as an absence of beta and gamma change. Beta change occurs in instances where the construct of interest remains constant, but the measurement properties of the indicators of the construct are temporally inconsistent (e.g., numerical values across assessment points are not on the same measurement scale). Gamma change occurs when the meaning of the construct changes over time (e.g., the number of factors that represent the construct vary across assessment waves). In applied longitudinal research, measurement invariance is often simply (implicitly) assumed and not examined. However, when measurement is not invariant over time, it is misleading to analyze and interpret the temporal change in observed measures or latent constructs; that is, change may be misinterpreted as alpha change when in fact the precision of measurement of the construct, or the construct itself, varies across time. Thus, the examination of measurement invariance should precede applications of SEM procedures with longitudinal data (e.g., latent growth curve models, autoregressive/cross-lagged panel models; Bollen & Curran, 2004; Curran & Hussong, 2003; Duncan, Duncan, Strycker, Li, & Alpert, 1999).

These procedures are illustrated using the longitudinal measurement model presented in Figure 7.2. In this example, the researcher wishes to evaluate whether an intervention was successful in improving employees' job satisfaction. Employees ($N = 250$) of a large company were administered four measures of job satisfaction (Measures A–D, which varied in assessment modality, e.g., questionnaires, supervisor ratings) immediately before and after the intervention (the pre- and posttest interval was 4 weeks). For each measure, higher scores reflect higher job satisfaction. Prior to examining whether the intervention resulted in an overall increase in job satisfaction, the researcher wishes to verify that the construct of job satisfaction, and its measurement, remain stable over time.

Figure 7.2 presents the hypothesized path model in which the construct of job satisfaction is posited to be structurally the same (i.e., unidimensional) at both assessment points (cf. gamma change). If the factor structure is temporally equivalent, additional tests can be performed to

TABLE 7.5. *(cont.)*

```
/LABELS
 v1=logical; v2=verbal; v3=word; v4=faces; v5=family; v6=visrep;
 f1 = auditory; f2 = visual;
/EQUATIONS
 V1 = *F1+E1;
 V2 = *F1+E2;
 V3 = *F1+E3;
 V4 = *F2+E4;
 V5 = *F2+E5;
 V6 = *F2+E6;
/VARIANCES
 F1 TO F2 = 1.0;
 E1 TO E6= *;
/COVARIANCES
 F1 TO F2 = *;
/CONSTRAINT
 (V1,F1) = (V2,F1) = (V3,F1);
 (V4,F2) = (V5,F2) = (V6,F2);
 (E1, E1) = (E2, E2) = (E3, E3);
 (E4, E4) = (E5, E5) = (E6, E6);
/MATRIX
<insert correlation matrix from Figure 7.1 here>
/STANDARD DEVIATIONS
2.61 2.66 2.59 1.94 2.03 2.05
/PRINT
 fit=all;
/LMTEST
/END
```

SAS 8.2 PROC CALIS

```
Title "CALIS SYNTAX FOR PARALLEL INDICATORS";
Data WMS (type=CORR);
 input _TYPE_ $ _NAME_ $ V1-V6;
 label V1 = 'logical'
 V2 = 'verbal'
 V3 = 'word'
 V4 = 'faces'
 V5 = 'family'
 V6 = 'visrep';
cards;
 mean .   0     0     0     0     0     0
  std . 2.61  2.66  2.59  1.94  2.03  2.05
   N  . 200   200   200   200   200   200
corr V1 1.000  .     .     .     .     .
corr V2 0.661 1.000  .     .     .     .
corr V3 0.630 0.643 1.000  .     .     .
corr V4 0.270 0.300 0.268 1.000  .     .
corr V5 0.297 0.265 0.225 0.805 1.000  .
corr V6 0.290 0.287 0.248 0.796 0.779 1.000
;
run;
```

(cont.)

TABLE 7.5. *(cont.)*

```
proc calis data=WMS cov method=ml pall pcoves;
var = V1-V6;
lineqs
 V1 = lam1 f1 + e1,
 V2 = lam1 f1 + e2,
 V3 = lam1 f1 + e3,
 V4 = lam2 f2 + e4,
 V5 = lam2 f2 + e5,
 V6 = lam2 f2 + e6;
std
 f1-f2 = 1.0,
 e1-e3 = td1,
 e4-e6 = td2;
cov
 f1-f2 = ph3;
run;
```

Amos Basic 5.0

```
' Example of Parallel Indicators in Amos 5.0

Sub Main ()
Dim sem As New AmosEngine

sem.TextOutput
sem.Standardized
sem.Smc

sem.BeginGroup "memory.txt"
sem.Structure "x1 <— AUDITORY (lam1)"
sem.Structure "x2 <— AUDITORY (lam1)"
sem.Structure "x3 <— AUDITORY (lam1)"
sem.Structure "x4 <— VISUAL (lam2)"
sem.Structure "x5 <— VISUAL (lam2)"
sem.Structure "x6 <— VISUAL (lam2)"
sem.Structure "x1 <— E1 (1)"
sem.Structure "x2 <— E2 (1)"
sem.Structure "x3 <— E3 (1)"
sem.Structure "x4 <— E4 (1)"
sem.Structure "x5 <— E5 (1)"
sem.Structure "x6 <— E6 (1)"
sem.Structure "E1 (err1)"
sem.Structure "E2 (err1)"
sem.Structure "E3 (err1)"
sem.Structure "E4 (err2)"
sem.Structure "E5 (err2)"
sem.Structure "E6 (err2)"
sem.Structure "AUDITORY (1)"
sem.Structure "VISUAL (1)"
sem.Structure "AUDITORY <—> VISUAL"

End Sub
```

TABLE 7.4. Selected Mplus Results of the Final Two-Factor Measurement Model of Memory

```
TESTS OF MODEL FIT

Chi-Square Test of Model Fit
          Value                          9.277
          Degrees of Freedom                16
          P-Value                       0.9016

CFI/TLI
          CFI                            1.000
          TLI                            1.009

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                       0.000
          90 Percent C.I.                0.000    0.028
          Probability RMSEA <= .05       0.989

SRMR (Standardized Root Mean Square Residual)
          Value                          0.027

MODEL RESULTS
              Estimates      S.E.   Est./S.E.      Std     StdYX
AUDITORY BY
X1                2.099     0.125     16.795     2.099     0.803
X2                2.099     0.125     16.795     2.099     0.803
X3                2.099     0.125     16.795     2.099     0.803

VISUAL BY
X4                1.782     0.097     18.364     1.782     0.890
X5                1.782     0.097     18.364     1.782     0.890
X6                1.782     0.097     18.364     1.782     0.890

VISUAL WITH
AUDITORY          0.381     0.070      5.431     0.381     0.381

Variances
AUDITORY          1.000     0.000      0.000     1.000     1.000
VISUAL            1.000     0.000      0.000     1.000     1.000

Residual Variances
X1                2.427     0.172     14.142     2.427     0.355
X2                2.427     0.172     14.142     2.427     0.355
X3                2.427     0.172     14.142     2.427     0.355
X4                0.832     0.059     14.142     0.832     0.208
X5                0.832     0.059     14.142     0.832     0.208
X6                0.832     0.059     14.142     0.832     0.208

R-SQUARE
Observed
Variable   R-Square
X1            0.645
X2            0.645
X3            0.645
X4            0.792
X5            0.792
X6            0.792
```


TABLE 7.5. Computer Syntax (LISREL, Mplus, EQS, CALIS, Amos) for Specification of Parallel Indicators of Auditory Memory and Visual Memory within a Two-Factor Measurement Model of Memory

LISREL 8.72

```
TITLE LISREL PROGRAM FOR PARALLEL INDICATORS
DA NI=6 NO=200 MA=CM
LA
X1 X2 X3 X4 X5 X6
KM
<insert correlation matrix from Figure 7.1 here>
SD
2.61 2.66 2.59 1.94 2.03 2.05
MO NX=6 NK=2 PH=SY,FR LX=FU,FR TD=DI
LK
AUDITORY VISUAL
PA LX
1 0
1 0
1 0
0 1
0 1
0 1
PA PH
0
1 0
VA 1.0 PH(1,1) PH(2,2)
EQ LX(1,1) LX(2,1) LX(3,1)
EQ LX(4,2) LX(5,2) LX(6,2)
EQ TD(1,1) TD(2,2) TD(3,3)
EQ TD(4,4) TD(5,5) TD(6,6)
OU ME=ML RS MI SC AD=OFF IT=100 ND=4
```

Mplus 3.11

```
TITLE:      MPLUS PROGRAM FOR PARALLEL INDICATORS
DATA:       FILE IS "C:\input6.dat";
            TYPE IS STD CORR;
            NOBS ARE 200;
VARIABLE:   NAMES ARE X1-X6;
ANALYSIS:   ESTIMATOR=ML;
MODEL:      AUDITORY BY X1* X2 X3 (1);
            VISUAL BY X4* X5 X6 (2);
            AUDITORY@1.0; VISUAL@1.0;
            X1 X2 X3 (3);
            X4 X5 X6 (4);
OUTPUT:     SAMPSTAT MODINDICES(4.00) STAND RESIDUAL;
```

EQS 5.7b

```
/TITLE
 EQS SYNTAX FOR PARALLEL INDICATORS
/SPECIFICATIONS
 CASES=200; VARIABLES=6; METHODS=ML; MATRIX=COR; ANALYSIS=COV;
```

(cont.)

TABLE 7.3. Model Fit of Congeneric, Tau-Equivalent, and Parallel Solutions of a Two-Factor Model of Memory (N = 200)

	χ^2	df	χ^2_{diff}	Δdf	RMSEA (90% CI)	CFit	SRMR	CFI	TLI
Congeneric solution	4.88	8			.000 (.000–.057)	.93	.012	1.00	1.01
Tau equivalent: Auditory Memory	5.66	10	0.78	2	.000 (.000–.044)	.96	.021	1.00	1.01
Tau equivalent: Auditory & Visual Memory	5.88	12	0.22	2	.000 (.000–.025)	.99	.022	1.00	1.01
Parallel: Auditory Memory	5.97	14	0.09	2	.000 (.000–.000)	1.00	.021	1.00	1.01
Parallel: Auditory & Visual Memory	9.28	16	3.31	2	.000 (.000–.028)	.99	.027	1.00	1.01

Note. χ^2_{diff}, nested χ^2 difference; RMSEA, root mean square error of approximation; 90% CI, 90% confidence interval for RMSEA; CFit, test of close fit (probability RMSEA ≤ .05); SRMR, standardized root mean square residual; CFI, comparative fit index; TLI, Tucker–Lewis Index. All χ^2 and χ^2_{diff} values nonsignificant (ps > .05).


```
MODEL RESULTS
                    Estimates      S.E.   Est./S.E.       Std     StdYX
AUDITORY BY
   X1                   2.100     0.125      16.800     2.100     0.807
   X2                   2.100     0.125      16.800     2.100     0.804
   X3                   2.100     0.125      16.800     2.100     0.798

VISUAL   BY
   X4                   1.756     0.108      16.183     1.756     0.907
   X5                   1.795     0.115      15.608     1.795     0.887
   X6                   1.796     0.117      15.378     1.796     0.878
```

Because this restriction of factor loading equality holds, it will be retained in subsequent tests of the two-factor measurement model of memory.

Next, the model is respecified with the added constraint of holding the factor loadings of Visual Memory to equality, to evaluate the tau equivalence of these indicators. As seen in Table 7.3, this restriction also does not significantly degrade the fit of the solution, $\chi^2_{diff}(2)$ = 0.22, *ns*; therefore, the indicators of Visual Memory can also be considered to be tau equivalent. Because tau equivalence has been established, the analysis can proceed to evaluating the condition of parallel indicators.

The current example defined the metric of Auditory Memory and Visual Memory by fixing their variances to 1.0. It is noteworthy that the same results would be obtained if the unstandardized factor loadings of all indicators were set to 1.0 (with the factor variances freely estimated). This is because when an indicator is selected to be a marker, its unstandardized loading is fixed to 1.0. Thus, if all factor loadings are to be tested for equality (tau equivalence), they must also equal the value of the marker (1.0).

In addition to tau equivalence, the condition of parallelism requires that the error variances of indicators loading on the same factor are the same. This added restriction is tested for Auditory Memory by placing equality constraints on the measurement errors of X1, X2, and X3. The findings presented in Table 7.3 show that this restriction does not result in a significant increase in χ^2, $\chi^2_{diff}(2)$ = 0.09, *ns*. Finally, the results in this table also indicate that the indicators of Visual Memory can be regarded as parallel, $\chi^2_{diff}(2)$ = 3.31, *ns*. The collective findings can be interpreted as being consistent with the notion that X1–X3 and X4–X6 are interchangeable indicators of the latent constructs of Auditory Memory and Visual Memory, respectively.

For the reader's information, the model estimates for the final solution are presented in Table 7.4. Table 7.5 provides the programming syntax for the final model in LISREL, Mplus, EQS, Amos, and CALIS. Before the dis-

TABLE 7.2. *(cont.)*

```
;
run;

proc calis data=WMS cov method=ml pall pcoves;
var = V1-V6;
lineqs
 V1 = lam1 f1 + e1,
 V2 = lam1 f1 + e2,
 V3 = lam1 f1 + e3,
 V4 = lam4 f2 + e4,
 V5 = lam5 f2 + e5,
 V6 = lam6 f2 + e6;
std
 f1-f2 = 1.0,
 e1-e6 = td1-td6;
cov
 f1-f2 = ph3;
run;
```

Amos Basic 5.0

```
' Example of Tau Equivalence in Amos 5.0

Sub Main ()
Dim sem As New AmosEngine

sem.TextOutput
sem.Standardized
sem.Smc

sem.BeginGroup "memory.txt"

sem.Structure "x1 <— AUDITORY (lam1)"
sem.Structure "x2 <— AUDITORY (lam1)"
sem.Structure "x3 <— AUDITORY (lam1)"
sem.Structure "x4 <— VISUAL"
sem.Structure "x5 <— VISUAL"
sem.Structure "x6 <— VISUAL"

sem.Structure "x1 <— E1 (1)"
sem.Structure "x2 <— E2 (1)"
sem.Structure "x3 <— E3 (1)"
sem.Structure "x4 <— E4 (1)"
sem.Structure "x5 <— E5 (1)"
sem.Structure "x6 <— E6 (1)"

sem.Structure "AUDITORY (1)"
sem.Structure "VISUAL (1)"
sem.Structure "AUDITORY <—> VISUAL"

End Sub
```


AUDITORY BY X1* X2 X3 (1);

where parameters are constrained to be equal by placing the same number in parentheses following the parameters that are to be held equal. An "*" is placed directly after the X1 indicator to override the Mplus default of setting the first indicator in the list as the marker indicator. This was done because the metric of Auditory Memory was defined by fixing its variance to 1.0. In EQS, the same constraints are imposed by adding a /CONSTRAINT section to the command file as follows:

/CONSTRAINT
(V1, F1) = (V2,F1) = (V3,F1);

In CALIS, these constraints are made in the following lines:

V1 = lam1 f1 + e1,
V2 = lam1 f1 + e2,
V3 = lam1 f1 + e3,

in which the factor loadings for the first factor (f1; Auditory Memory) are each given the same parameter name (lam1; the user may provide any name for these or other model parameters). The same strategy is used in Amos Basic (parameters held to equality given same label, lam1).

As shown in Table 7.3, this restriction results in an increase in χ^2 to 5.66 with $df = 10$ ($p = .84$); for a more in-depth discussion of model estimation under this equality constraint, see Appendix 7.1. This solution has 2 more *df*s than the congeneric model ($df = 8$) because of the equality constraint on the factors loadings of Auditory Memory; that is, the congeneric model contains three separate estimates of the X1, X2, and X3 factor loadings, and the tau-equivalent model contains one factor loading estimate applied to each of these three indicators. The difference in χ^2 of the current solution and the congeneric solution is 0.78, with $df = 2$. Because this χ^2 difference is below the critical value of the χ^2 distribution at $df = 2$ (i.e., $\chi^2_{\text{crit}} = 5.99$, at $\alpha = .05$), it can be concluded that the three indicators of Auditory Memory are tau equivalent; that is, a unit increase in the latent construct of Auditory Memory is associated with the same amount of change in each of the X1, X2, and X3 indicators. The selected Mplus output provided below shows that these equality constraints apply to both the unstandardized factor loadings and their standard errors (and hence the *z* tests of significance as well).

TABLE 7.2. Computer Syntax (LISREL, Mplus, EQS, CALIS, Amos) for Specification of Tau-Equivalent Indicators of Auditory Memory within a Two-Factor Measurement Model of Memory

LISREL 8.72

```
TITLE LISREL PROGRAM FOR TAU EQUIVALENT AUDITORY INDICATORS
DA NI=6 NO=200 MA=CM
LA
X1 X2 X3 X4 X5 X6
KM
1.000
0.661 1.000
0.630 0.643 1.000
0.270 0.300 0.268 1.000
0.297 0.265 0.225 0.805 1.000
0.290 0.287 0.248 0.796 0.779 1.000
SD
2.61 2.66 2.59 1.94 2.03 2.05
MO NX=6 NK=2 PH=SY,FR LX=FU,FR TD=DI
LK
AUDITORY VISUAL
PA LX
1 0
1 0
1 0
0 1
0 1
0 1
PA PH
0
1 0
VA 1.0 PH(1,1) PH(2,2)
EQ LX(1,1) LX(2,1) LX(3,1)              ! EQUALITY CONSTRAINT
OU ME=ML RS MI SC AD=OFF IT=100 ND=4
```

Mplus 3.11

```
TITLE:     MPLUS PROGRAM FOR TAU EQUIVALENT AUDITORY INDICATORS
DATA:      FILE IS "C:\input6.dat";
           TYPE IS STD CORR;
           NOBS ARE 200;
VARIABLE:  NAMES ARE X1-X6;
ANALYSIS:  ESTIMATOR=ML;
MODEL:     AUDITORY BY X1* X2 X3 (1);  ! EQUALITY CONSTRAINT
           VISUAL BY X4* X5 X6;
           AUDITORY@1.0; VISUAL@1.0;
OUTPUT:    SAMPSTAT MODINDICES(4.00) STAND RESIDUAL;
```

EQS 5.7b

```
/TITLE
 EQS SYNTAX FOR TAU EQUIVALENT AUDITORY INDICATORS
/SPECIFICATIONS
 CASES=200; VARIABLES=6; METHODS=ML; MATRIX=COR; ANALYSIS=COV;
```

(cont.)

TABLE 7.2. *(cont.)*

```
/LABELS
 v1=logical; v2=verbal; v3=word; v4=faces; v5=family; v6=visrep;
 f1 = auditory; f2 = visual;
/EQUATIONS
 V1 = *F1+E1;
 V2 = *F1+E2;
 V3 = *F1+E3;
 V4 = *F2+E4;
 V5 = *F2+E5;
 V6 = *F2+E6;
/VARIANCES
 F1 TO F2 = 1.0;
 E1 TO E6= *;
/COVARIANCES
 F1 TO F2 = *;
/CONSTRAINT
 (V1,F1) = (V2,F1) = (V3,F1);
/MATRIX
1.000
0.661 1.000
0.630 0.643 1.000
0.270 0.300 0.268 1.000
0.297 0.265 0.225 0.805 1.000
0.290 0.287 0.248 0.796 0.779 1.000
/STANDARD DEVIATIONS
2.61 2.66 2.59 1.94 2.03 2.05
/PRINT
 fit=all;
/LMTEST
/END
```

SAS 8.2 PROC CALIS

```
Title "CALIS SYNTAX FOR TAU EQUIVALENT AUDITORY INDICATORS";
Data WMS (type=CORR);
 input _TYPE_ $ _NAME_ $ V1-V6;
 label V1 = 'logical'
 V2 = 'verbal'
 V3 = 'word'
 V4 = 'faces'
 V5 = 'family'
 V6 = 'visrep';
cards;
mean  .   0      0      0      0      0      0
 std  .   2.61   2.66   2.59   1.94   2.03   2.05
   N  .   200    200    200    200    200    200
corr V1   1.000  .      .      .      .      .
corr V2   0.661  1.000  .      .      .      .
corr V3   0.630  0.643  1.000  .      .      .
corr V4   0.270  0.300  0.268  1.000  .      .
corr V5   0.297  0.265  0.225  0.805  1.000  .
corr V6   0.290  0.287  0.248  0.796  0.779  1.000
```

(cont.)

using the less restricted congeneric model as the baseline solution. If the results are in accord with tau equivalence, that is, constraining the factor loadings to equality does not result in a significant increase in χ^2, the analysis can proceed to the evaluation of parallel indicators.

To illustrate this procedure, a two-factor model of memory is used in which three observed measures are posited to be indicators of the construct of Auditory Memory (X1 = logical memory, X2 = verbal paired association, X3 = word list) and three additional measures are posited to be indicators of the construct of Visual Memory (X4 = faces, X5 = family pictures, X6 = visual reproduction) (the example is loosely based on the structure of the Wechsler Memory Scale, Wechsler, 1997; for applied CFA examples in this research domain, see Price, Tulsky, Millis, & Weiss, 2002, and Tulsky & Price, 2003). The measurement model is presented in Figure 7.1 along with the sample (N = 200) correlations and standard deviations of the six indicators.

The congeneric, two-factor solution provides a good fit to the data, $\chi^2(8) = 4.88$, $p = .77$, SRMR = .012, RMSEA = 0.00 (90% CI = 0.00 to 0.06, CFit = .93), TLI = 1.01, CFI = 1.00. The inspection of modification indices and standardized residuals reveals no areas of strain in the solution. The model parameter estimates are presented in Table 7.1. All six factor loadings are statistically significant ($ps < .001$) and sufficiently large (range of completely standardized estimates = .78 to .91) (in this example, the metric of the latent variables was defined by fixing factor variances to 1.0). As expected, the constructs of Auditory Memory and Visual Memory are significantly correlated ($p < .001$), $\phi_{21} = .38$.

First, the tau equivalence of the indicators that load on Auditory Memory are examined. This is performed by placing an equality constraint on the factor loadings of the X1, X2, and X3 indicators. Table 7.2 provides the programming syntax for this model in the LISREL, Mplus, EQS, Amos, and CALIS languages. As shown in this table, the software programs differ in their syntax for imposing equality constraints on the model parameters. In LISREL, this restriction is accomplished by adding the line

EQ LX(1,1) LX(2,1) LX(3,1)

where EQ is the keyword for holding the parameters that follow to equality. The syntax also illustrates a programming shortcut in LISREL, TD = DI, where the theta-delta matrix is specified to be diagonal (DI), meaning that the indicator error variances are freely estimated and the error covariances are fixed to 0 (i.e., the only freed parameters within TD are on the diagonal). In Mplus, the equality constraints are reflected in the line

TABLE 7.1. Mplus Results of the Two-Factor Model of Memory

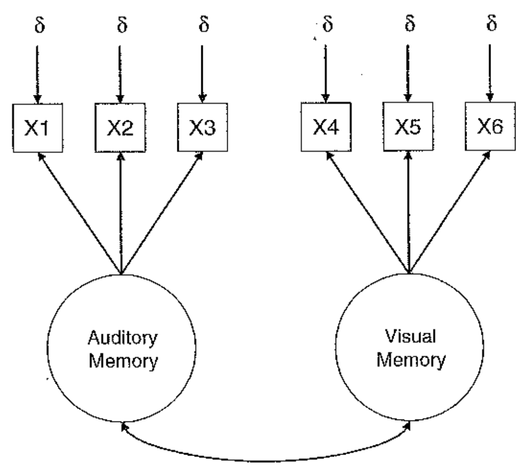
```
TESTS OF MODEL FIT

Chi-Square Test of Model Fit
          Value                              4.877
          Degrees of Freedom                     8
          P-Value                           0.7706

CFI/TLI
          CFI                                1.000
          TLI                                1.008

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.000
          90 Percent C.I.                    0.000   0.057
          Probability RMSEA <= .05           0.929

SRMR (Standardized Root Mean Square Residual)
          Value                              0.012

MODEL RESULTS
                Estimates      S.E.   Est./S.E.      Std      StdYX
AUDITORY BY
 X1                 2.101     0.166     12.663     2.101      0.807
 X2                 2.182     0.168     12.976     2.182      0.823
 X3                 2.013     0.166     12.124     2.013      0.779

VISUAL   BY
 X4                 1.756     0.108     16.183     1.756      0.907
 X5                 1.795     0.115     15.608     1.795      0.887
 X6                 1.796     0.117     15.378     1.796      0.878

VISUAL   WITH
 AUDITORY           0.382     0.070      5.464     0.382      0.382

Variances
 AUDITORY           1.000     0.000      0.000     1.000      1.000
 VISUAL             1.000     0.000      0.000     1.000      1.000

Residual Variances
 X1                 2.366     0.372      6.365     2.366      0.349
 X2                 2.277     0.383      5.940     2.277      0.323
 X3                 2.620     0.373      7.027     2.620      0.393
 X4                 0.662     0.117      5.668     0.662      0.177
 X5                 0.877     0.134      6.554     0.877      0.214
 X6                 0.956     0.139      6.866     0.956      0.229

R-SQUARE
Observed
Variable   R-Square
X1           0.651
X2           0.677
X3           0.607
X4           0.823
X5           0.786
X6           0.771
```

the corresponding unconstrained solution, it can be concluded that the four indicators have equivalent relationships to the latent factor.

EQUALITY CONSTRAINTS WITHIN A SINGLE GROUP

Congeneric, Tau-Equivalent, and Parallel Indicators

In Chapter 3, it was noted that most CFA specifications in applied research entail congeneric indicator sets. Congeneric indicators are presumed to measure the same construct, and the size of their factor loadings and measurement errors are free to vary; in addition, the assumption of independent measurement errors must hold. Figure 7.1 provides an example of a



Sample Correlations and Standard Deviations (SDs); N = 200

	X1	X2	X3	X4	X5	X6
X1	1.000					
X2	0.661	1.000				
X3	0.630	0.643	1.000			
X4	0.270	0.300	0.268	1.000		
X5	0.297	0.265	0.225	0.805	1.000	
X6	0.290	0.287	0.248	0.796	0.779	1.000
SD:	2.610	2.660	2.590	1.940	2.030	2.050

FIGURE 7.1. Two-factor model of memory. X1, Logical Memory; X2, Verbal Paired Association; X3, Word List; X4, Faces; X5, Family Pictures; X6, Visual Reproduction.

congeneric model in which the first three observed measures (X1, X2, X3) are indicators of one latent construct (Auditory Memory), and the second set of measures (X4, X5, X6) are indicators of another latent construct (Visual Memory). In addition to congeneric models, the psychometric literature distinguishes more restrictive solutions that test for the conditions of *tau-equivalent* and *parallel* indicators. A tau-equivalent model entails a congeneric solution in which the indicators of a given factor have equal loadings but differing error variances. As noted previously, when the condition of equal factor loadings holds, it can be asserted that the indicators have equivalent relationships with the underlying construct they measure. Stated another way, a one-unit change in the latent variable is associated with the same amount of change in each indicator that loads on that factor. The most restrictive solution treats indicators as parallel, in which the observed measures are posited to have equal factor loadings and equal error variances. Thus, in addition to the assumption that indicators measure the latent construct in the same units of measurement (tau equivalence), parallel indicators are assumed to measure the latent construct with the same level of precision (i.e., reflected by equivalent error variances).[1] These distinctions have psychometric implications. For instance, Raykov (2001a) has shown that Cronbach's coefficient alpha is a misestimator of the scale reliability of a multicomponent measuring instrument (e.g., multiple-item questionnaire) when the assumption of tau equivalence does not hold (in Chapter 8, a CFA-based approach to estimating scale reliability is presented that does not rest on this assumption). If the conditions of parallel indicators are met, this lends support for the notion that the measures in question are psychometrically interchangeable, a finding that is germane to the endeavor of establishing parallel test forms or justifying the practice of operationalizing a latent construct by summation of its indicators' observed scores (McDonald, 1999).

The test for tau-equivalent and parallel indicators begins with the evaluation of the congeneric measurement model in which the factor loadings and residual variances are free to vary. If the conditions of a congeneric solution are not met (e.g., an indicator loads on more than one factor), the analysis would not proceed to the evaluation of tau equivalence unless, perhaps, substantive considerations allow for the measurement model to be revised to conform to a congeneric model (e.g., removal of a cross-loading indicator). In the context of a congeneric measurement model, the test for tau equivalence is conducted by placing the appropriate restrictions on the solution (i.e., equality constraints on indicators that load on the same factor) and evaluating the resulting change in model χ^2

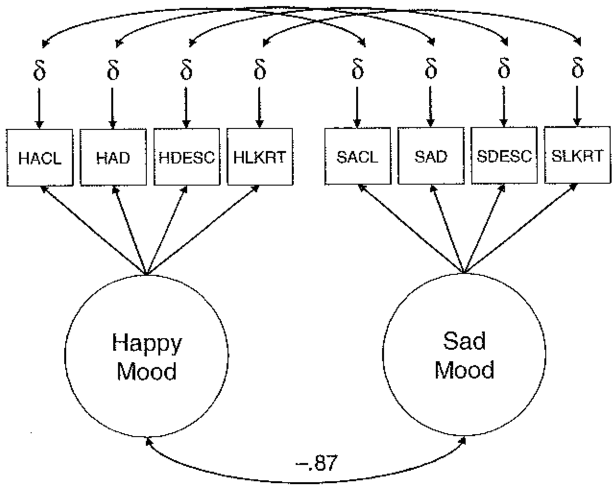
7

CFA with Equality Constraints, Multiple Groups, and Mean Structures

The previous CFA examples presented in this book have been estimated within a single group, used a variance–covariance matrix as input, and entailed model parameters that were either freely estimated or fixed. In this chapter, these analyses are extended in several ways. For instance, some CFA specifications will place equality constraints on selected parameters of the measurement model. Such constraints may be applicable in CFA analyses involving a single group (e.g., do the items of a questionnaire assess the same latent construct in equivalent units of measurement?) or two or more groups (e.g., do males and females respond to items of a measuring instrument in a similar manner?). In addition, two different approaches to CFA with multiple groups are presented (multiple-groups CFA, MIMIC models) in context of the analysis of measurement invariance and population heterogeneity. Finally, CFA is extended to the analysis of mean structures involving the estimation of indicator intercepts and latent factor means and the evaluation of their equivalence in multiple groups. Consequently, the input matrix must include the sample means of the indicators, in addition to their variances and covariances. The substantive applications of each approach are discussed.

OVERVIEW OF EQUALITY CONSTRAINTS

As first noted in Chapter 3, parameters in a CFA solution can be freely estimated, fixed, or constrained. A free parameter is unknown, and the researcher allows the analysis to find its optimal value that, in tandem with other model estimates, minimizes the differences between the observed and predicted variance–covariance matrices (e.g., in a one-factor CFA model, obtain the set of factor loadings that best reproduce the observed correlations among four input indicators). A fixed parameter is prespecified by the researcher to be a specific value, most commonly either 1.0 (e.g., in the case of marker indicators or factor variances to define the metric of a latent variable) or 0 (e.g., the absence of cross-loadings or error covariances). As with a free parameter, a constrained parameter is also unknown. However, the parameter is not free to be any value, but rather the specification places restrictions on the values it may assume. The most common form of constrained parameters are *equality constraints*, in which unstandardized parameters are restricted to be equal in value (i.e., see the section in Chapter 8 on scale reliability evaluation for a different type of constrained parameter). Consider a CFA model in which four indicators are specified to load on a single factor. If the CFA specification entails an equality constraint on the four factor loadings, the specific value of these loadings is unknown a priori (as in models with freely estimated factor loadings), but the analysis must find a single estimate (applied to all four loadings) that best reproduces the observed relationships among the four indicators. This is unlike the models with freely estimated parameters where the factor loadings are free to take on any set of values that maximize the fit of the solution. Two important principles apply to the various examples of equality constraints discussed in this chapter. First, as with fixed parameters, these constraints are placed on the unstandardized solution. Accordingly, the indicators whose parameters are to be held equal should have the same metric; for example, the unstandardized loading of an indicator defined by a 0–100 scale will inherently differ from the unstandardized loading of an indicator measured on a 1–8 scale. Second, because a CFA model with equality constraints is nested under the measurement model without these constraints (i.e., it entails a subset of the parent model's freely estimated parameters), χ^2 difference testing can be employed as a statistical comparative evaluation of the constrained solution. For example, if the CFA model where the four factor loadings are held to equality does not produce a significant reduction in fit relative to



Correlations (N = 304):

	HACL	HAD	HDESC	HLKRT	SACL	SAD	SDESC	SLKRT
HACL	1.00							
HAD	.60	1.00						
HDESC	.59	.69	1.00					
HLKRT	.59	.67	.74	1.00				
SACL	-.10	-.46	-.53	-.50	1.00			
SAD	-.44	-.53	-.63	-.64	.68	1.00		
SDESC	-.44	-.48	-.59	-.59	.52	.60	1.00	
SLKRT	-.47	-.59	-.65	-.66	.60	.69	.61	1.00

FIGURE 6.3. Path diagram and input data for CFA model of positive and negative mood. Correlation matrix is taken from Green et al. (1993, Table 7). The first letter of each indicator refers to the construct it purportedly measures: H, happy; S, sad. The remaining letters of each indicator pertain to the self-report assessment format: ACL, adjective checklist; AD, agree–disagree format; DESC, self-descriptive format; LKRT, semantic differential Likert scale.

ture. This is presumably due in part to the fact that, in many cases, multiple measures are not available for a given construct (e.g., the correlated uniqueness approach ordinarily requires at least three assessment methods). Although the examples used in this chapter involved relatively disparate assessment methods, these approaches can be employed for constructs assessed by a single assessment modality; for example, Cluster A personality dimensions assessed by three different self-report inventories (also see Green et al., 1993).

MTMM models can be extended in useful ways. First, the MTMM model could be embedded in a larger structural equation model, such as one that relates the trait factors to background variables or distal outcomes that represent external validators; for example, are the Cluster A personality dimensions differentially related to salient clinical variables such as comorbidity, overall functioning, or long-term adjustment? Such analyses could strongly bolster the importance of the substantive constructs of the MTMM model (e.g., establish their predictive validity). A second possible extension is analyzing the MTMM model within a multiple-groups solution to determine whether salient parameters (e.g., trait factor loadings) are invariant across salient subgroups (e.g., is each observed measure of Cluster A personality related to the substantive trait in an equivalent manner for men and women?). Approaches to evaluating invariance of CFA models are presented in Chapter 7, along with the analysis of mean structures (i.e., indicator intercepts, latent factor means).

NOTES

1. For instance, Tomás et al. (2000) found that, under some conditions, the correlated methods model works reasonably well (and perhaps better than the correlated uniqueness approach) when more than two indicators per trait–method combination are available. However, because the typical MTMM study is a $3T \times 3M$ design with a single indicator per each trait–method combination (Marsh & Grayson, 1995), evaluation of the correlated methods model under the conditions studied by Tomás et al. (2000) may not be practical in many applied research scenarios.

2. The CT–C(M–1) model can be identified with as few as two traits and two methods. However, to separate trait, method, and error effects in the model, at least two indicators per each trait–method combination are required (e.g., for each trait in Figure 6.1, at least two inventory indicators, two clinical interview indicators, and two observer rating indicators are needed; p = 18).

3. In addition, a comprehensive review of the most common types of method effects found in social and behavioral sciences research is provided by Podsakoff et al. (2003).

acquiescent response style). In addition, this specification yields better estimates of the relationships between the indicators and the latent construct (i.e., reduces bias in the factor loadings).

These points also apply to the CFA of MTMM data. For instance, when the data in Table 6.1 are reanalyzed without modeling method variance (i.e., three trait factors are specified, all measurement error is presumed to be random), a poorly fitting solution results, χ^2 (24) = 454.41, p < .001, SRMR = .063, RMSEA = 0.183 (90% CI = 0.168 to 0.199, CFit < .001), TLI = .765, CFI = .843. Inspection of standardized residuals and modification indices reveal that relationships among indicators obtained from the same assessment method are not well explained by the model; for example, standardized residuals (LISREL) for observer ratings:

	PARO	SZTO	SZDO
PARO	- -		
SZTO	4.3866	- -	
SZDO	3.9154	3.6211	- -

The failure to account for method variance may lead to the false conclusion that the constructs under study have poor discriminant validity. This is because the correlations among the factors (traits) are apt to be inflated by the CFA estimation process. Because method effects have not been taken into account in the model specification, the estimation process attempts to reproduce the additional covariance of indicators sharing the same assessment method by increasing the magnitude of the factor correlations (cf. Eq. 3.8, Chapter 3). This is exemplified in the current data set, where each of the correlations among the Cluster A personality traits are inflated when method effects are not taken into account; for example, correlation of Schizotypal and Schizoid = .31 and .36 in CFA solutions with and without method effects, respectively. This is less problematic in the Table 6.1 data since the parameter estimates would not be interpreted in the context of a poorly fitting model (e.g., RMSEA = 0.183), and the factor correlations in both solutions are not large (e.g., .31 vs. .36). However, this will not always be the case in applied data sets (e.g., inflation of factor correlations can be more extreme).

Another good example in the applied literature on the effects of (not) modeling measurement error is in the area of emotion. Investigators have long debated whether positive and negative mood (e.g., happiness and sadness) are largely independent (i.e., separate constructs) or bipolar (i.e., represent opposite ends of a single dimension). Although original theories of emotion assumed bipolarity, much of the early research did not support this position. For instance, the initial evidence showed that single indicators of happy and sad moods were not strongly correlated. Consequently, EFAs typically revealed two-factor solutions, along the lines of EFAs of questionnaires containing reversed items (described above). However, researchers later came to learn that the low correlations between indicators of positive and negative moods may be due to the failure to consider and model the effects of random and nonrandom response error (e.g., Green, Goldman, & Salovey, 1993; Russell, 1979). A thorough description of the sources of systematic error in the measurement of affect can be found in Russell (1979) and Green et al. (1993).[3] But when random and systematic error is adjusted for, the correlation between positive and negative emotions increases to an extent that severely challenges the differentiation (discriminant validity) of these dimensions.

To illustrate, a sample correlation matrix from Green et al. (1993) is presented in Figure 6.3 (N = 304 undergraduates). The emotions of happiness and sadness were assessed by four different types of self-report formats: adjective checklist, item response options ranging from strong agreement to strong disagreement, self-description response options ranging from very well to not at all, and a semantic differential Likert scale. Some of the correlations seen in Figure 6.3 might be mistakenly interpreted in support of the notion that happiness and sadness are distinct dimensions; for example, the correlation between adjective checklist indicators of happiness and sadness is –.10. When a CFA model with the appropriate error theory is fit to these data (see path diagram in Figure 6.3), a different picture emerges. The two-factor measurement model, which takes random and nonrandom error into account, provides a reasonable fit to the data, $\chi^2(15)$ = 26.13, p = .036, SRMR = .026, RMSEA = 0.049 (90% CI = 0.01 to 0.08, CFit = .487), TLI = .993, CFI = .996. However, the factor correlation of Happiness and Sadness is found to be –.87, a result that seriously challenges the discriminant validity of these constructs and upholds the contention of the bipolarity of emotion. These findings strongly underscore the importance of multimethod research designs and analytic approaches (i.e., CFA) that model measurement error in construct validation.

SUMMARY

CFA approaches to MTMM matrices (and the MTMM approach in general) continue to be somewhat underutilized in many areas of the applied litera-

1990). Unlike the preceding CFA approaches, the direct product model addresses the possibility that method factors interact with trait factors in a multiplicative manner rather than additively (cf. Campbell & O'Connell, 1967). In other words, method effects may augment the correlations of strongly related traits more than they augment the correlations of weakly related constructs; for example, the higher the correlation between traits, the greater the effects of methods. When the data conform to the direct product model, this parameterization provides an elegant test of the criteria outlined by Campbell and Fiske (1959) for evaluating convergent and discriminant validity (for applied examples, see Bagozzi & Yi, 1990; Coovert, Craiger, & Teachout, 1997; and Lievens & Conway, 2001). Another advantage of this method is that it estimates a correlation matrix for the methods. This matrix can be inspected to evaluate the similarity of methods; for example, high correlations between purportedly distinct methods would challenge their discriminant validity (i.e., the methods are in effect the same). However, the direct product model is relatively difficult to program and interpret (the reader is referred to Bagozzi & Yi, 1990, and Wothke, 1996, for fully worked through LISREL parameterizations of direct product solutions). The direct product model often produces improper solutions, although much less so than correlated methods approaches (e.g., Lievens & Conway, 2001). A number of other disadvantages have been noted. For instance, some researchers (e.g., Podsakoff et al., 2003) have concluded that the extant evidence indicates that trait × method interactions are not very common or potent and thus a simpler CFA parameterization will usually be just as good. Nevertheless, more extensive evaluation of the direct product model is needed.

A newer alternative is the *correlated trait–correlated method minus one model*, CT–C(*M*–1) (Eid, 2000; Eid, Lischetzke, Nussbeck, & Trierweiler, 2003). The CT-C(*M*-1) model is very similar to the correlated methods model (Figure 6.1), except that it contains one method factor less than the methods included (*M*–1); for example, the path diagram in Figure 6.1 could be converted into a CT–C(*M*–1) model by eliminating the Inventory method factor.[2] This parameterization resolves some identification problems of the correlated methods model. The omitted method factor becomes the "comparison standard." The principal notion behind the CT–C(*M*–1) model is that, for each trait, the true-score variables of indicators of the comparison standard are regressors in a latent regression analysis in which the dependent measures are the true-score variables of the "nonstandard" methods ("nonstandard" indicates methods other than the comparison standard). Thus, a method factor is a residual factor common to all variables measured by the same method. In other words, it represents the portion of a trait measured by a nonstandard method that cannot be predicted by the true-score variable of the indicators measured by the comparison standard method. A step-by-step illustration of the CT–C(*M*–1) model is presented in Eid et al. (2003); for an applied example, see Lischetzke and Eid (2003). The CT–C(*M*–1) has many of the strengths of the correlated methods model (e.g., straightforward decomposition of trait, method, and error effects) but avoids the serious problems of underidentification and improper solutions. The CT–C(*M*–1) model has several other advantages. When multiple indicators are used (see note 2), the CT–C(*M*–1) model can test for trait-specific method effects (i.e., there exists a source of variance specific to each trait–method combination), allowing the researcher to examine the generalizability of method effects across traits. Moreover, the model also provides correlations among methods factors, and to some extent, information on the relationships between method factors and trait factors. One limitation is the asymmetry of the CT–C(*M*–1) model because a comparison standard method must be selected. Choice of the comparison standard may be clear-cut in some situations (e.g., when a "gold standard" exists), but not others (e.g., when all methods are somewhat similar or randomly chosen). Like the direct product model, the performance of the CT–C(*M*–1) model in applied and simulated data sets has yet to receive extensive study.

CONSEQUENCES OF NOT MODELING METHOD VARIANCE AND MEASUREMENT ERROR

It was noted in earlier chapters (e.g., Chapters 3 and 5) that the ability to model method effects is an important advantage of CFA over EFA. For instance, EFAs of questionnaires that were designed to measure unidimensional constructs (e.g., self-esteem) tend to produce two-factor solutions (Positive Self-Esteem, Negative Self-Esteem). These multidimensional latent structures stem from method effects introduced by including reverse-worded items. A conceptually more viable unidimensional structure can be upheld by CFA through specification of a single trait factor (e.g., Self-Esteem) and method effects (e.g., correlated errors) among the reverse-worded items (e.g., Marsh, 1996). In CFA, specification of a one-factor model without method effects would result in a poor-fitting solution. The correlated errors are needed to account for the additional covariance among items that is due to nonrandom measurement error (e.g.,

(1992) have shown that research data frequently approximate these cases (e.g., factor loading estimates that are roughly equal in magnitude). When factor loadings are roughly equal or if discriminant validity is poor, this empirical underidentification results in severe estimation difficulties. Because of these problems, Kenny and Kashy (1992) and other methodologists (e.g., Marsh & Grayson, 1995) have recommended the use of the correlated uniqueness model over the correlated methods approach for the analysis of MTMM data. However, other researchers (Lance, Noble, & Scullen, 2002) have concluded that, given the substantive strengths of the correlated methods model, the correlated uniqueness model should be used only if the correlated methods model fails. These authors underscore the various design features that may foster the chances of obtaining an admissible correlated methods solution (e.g., increased sample size, larger MTMM designs).[1]

Although their propensity for estimation problems is the primary disadvantage of correlated methods models, researchers have noted another limitation with this specification approach. In particular, these models do not allow for multidimensional method effects. A method effect would be multidimensional if there exist two or more systematic sources of variability (aside from the underlying trait) that affect some or all of the indicators in the model (e.g., response set, reverse-worded questionnaire items, indicators differentially affected by social desirability or the tendency for over- or underreporting). As noted earlier, correlated methods models attempt to explain all the covariance associated with a given assessment method by a single method factor; hence, the methods effects are assumed to be unidimensional. Although these models do not permit correlations between trait and method factors, researchers have demonstrated instances where this assumption is unrealistic (e.g., Kumar & Dillon, 1992). Moreover, methodologists have illustrated scenarios in which the partitioning of observed measure variance into trait and method components does not produce trait-free and method-free interpretations (e.g., Bagozzi, 1993).

Unlike the correlated methods model, correlated uniqueness models rarely pose estimation problems. For instance, in the Marsh and Bailey (1991) study, correlated uniqueness model specifications resulted in proper solutions 98% of the time. In addition, these models can accommodate both unidimensional and multidimensional method effects because method covariance is reproduced by freely estimated correlations (which may differ greatly in magnitude) among indicators based on the same assessment approach. However, the interpretation of correlated uniquenesses as method effects is not always straightforward (e.g., Bagozzi, 1993). Although this parameterization allows for multidimensional method effects, the resulting solution does not provide information on the interpretative nature of these effects (e.g., does not foster a substantive explanation for why the magnitude of the correlated errors may vary widely), and so, for instance, it can be difficult to determine which method has the most method variance.

Another potential drawback of the correlated uniqueness model is its assumption that the correlations among methods, and the correlations between traits and methods, are zero. In fact, when the number of traits (T) is three, the correlated uniqueness model and the uncorrelated methods model virtually always produce equivalent solutions (e.g., model *df* and goodness of fit are identical; the parameter estimates from one solution can be transformed into the other). Methodologists (e.g., Byrne & Goffin, 1993; Kenny & Kashy, 1992) have shown how the parameter estimates of a correlated uniqueness solution may be biased when the assumption of zero correlations among methods and between methods and traits does not hold. If these zero correlation constraints are not tenable, the amount of trait variance and covariance between trait factors will be overestimated, resulting in inflated estimates of convergent validity and lower estimates of discriminant validity, respectively (Kenny & Kashy, 1992; Marsh & Bailey, 1991). Kenny and Kashy (1992) have illustrated this biasing effect in situations where the covariances among methods are mistakenly assumed to be zero and the true factor loadings are all equal:

> The average method–method covariance is added to each element of the trait–trait covariance matrix. So, if the methods are similar to one another, resulting in positive method–method covariances, the amount of trait variance will be overestimated as will be the amount of trait–trait covariance. (pp. 169–170)

Although the size of these biases is usually trivial (Marsh & Bailey, 1991), Kenny and Kashy (1992) nonetheless recommend that researchers using correlated uniqueness models to analyze MTMM data should try to employ assessment methods that are as independent as possible.

OTHER CFA PARAMETERIZATIONS OF MTMM DATA

A number of other CFA-based strategies for analyzing MTMM data have been developed. One of the more prominent alternative approaches is the *direct product model* (Browne, 1984a; Cudeck, 1988; Wothke & Browne,

TABLE 6.4. Completely Standardized Estimates for the Correlated Uniqueness Model of the MTMM Matrix of Cluster A Personality Disorders

	Trait Factor Loadings			SMC	Uniqueness	Correlated Uniquenesses		
	Paranoid	Schizotypal	Schizoid					
PAR_T	.712			.507	.493	1.000		
SZT_T		.788		.621	.379	.094	1.000	
SZD_T			.769	.592	.408	.198	.293	1.000
PAR_C	.841			.707	.293	1.000		
SZT_C		.768		.589	.411	-.037	1.000	
SZD_C			.860	.740	.260	-.073	-.108	1.000
PAR_O	.788			.620	.380	1.000		
SZT_O		.843		.711	.289	.137	1.000	
SZD_O			.872	.760	.240	.111	.126	1.000

	Trait Factor Correlations		
	Paranoid	Schizotypal	Schizoid
Paranoid	1.000		
Schizotypal	.381	1.000	
Schizoid	.359	.310	1.000

Note. SMC, squared multiple correlation (i.e., λ^2). All unstandardized parameters are significantly different ($p < .05$) from zero except for the correlation between the unique variances of SZT_C and PAR_C.

ADVANTAGES AND DISADVANTAGES OF CORRELATED METHODS AND CORRELATED UNIQUENESS MODELS

As Kenny and Kashy (1992) note, an appealing feature of the correlated methods model is that it corresponds directly to Campbell and Fiske's (1959) original conceptualization of the MTMM matrix. Under this specification, each indicator is considered to be a function of trait, method, and unique variance. For instance, using the completely standardized solution, the squared trait factor loading (squared multiple correlations or communalities), squared method factor loading, and uniqueness for any given indicator sum to 1.0. Thus, these estimates can be interpreted as the proportions of trait, method, and unique variance of each indicator. Moreover, the parameter estimates produced by correlated methods solutions provide seemingly straightforward interpretations with regard to construct validity; for example, large trait factor loadings suggest favorable convergent validity, small or nonsignificant method factor loadings imply an absence of method effects, and modest trait factor intercorrelations suggest favorable discriminant validity. The specification of method factors fosters the substantive interpretation of method effects. Because the covariance associated with a given method is (it is hoped) accounted for by a single latent factor (e.g., a method factor for Observer Ratings; see Figure 6.1), method effects are assumed to be unidimensional (although, as discussed later, this assumption does not always hold). Moreover, unlike the correlated uniqueness model, the correlated methods approach allows for the evaluation of the extent to which method factors are intercorrelated.

However, an overriding drawback of the correlated methods model is that it is usually empirically underidentified. Consequently, a correlated methods solution will typically fail to converge. If it does converge, the solution will usually be associated with Heywood cases (Chapter 5) and large standard errors. For instance, using more than 400 MTMM matrices derived from real and simulated data, Marsh and Bailey (1991) found that the correlated methods model resulted in improper solutions 77% of the time. These authors noted that improper solutions were most probable when the MTMM design was small (e.g., $3T \times 3M$), when sample size was small, and when the assumption of unidimensional method effects was untenable. Moreover, Kenny and Kashy (1992) demonstrated that the correlated methods model was empirically underidentified in two special cases: (1) when the loadings on a trait or methods factor are equal; and (2) when there is no discriminant validity between two or more factors. Although these conditions are never perfectly realized, Kenny and Kashy

TABLE 6.3. *(cont.)*

```
/STANDARD DEVIATIONS
3.61 3.66 3.59 2.94 3.03 2.85 2.22 2.42 2.04
/PRINT
 fit=all;
/LMTEST
/WTEST
/END
```

SAS 8.2 PROC CALIS

```
Title "CALIS SYNTAX FOR CORRELATED UNIQUENESS MTMM SPECIFICATION";
Data CLUSTA (type=CORR);
 input _TYPE_ $ _NAME_ $ V1-V9;
 label V1 = 'pari'
       V2 = 'szti'
       V3 = 'szdi'
       V4 = 'parc'
       V5 = 'sztc'
       V6 = 'szdc'
       V7 = 'paro'
       V8 = 'szto'
       V9 = 'szdo';
cards;
mean  .      0      0      0      0      0      0      0      0      0
 std  .   3.61   3.66   3.59   2.94   3.03   2.85   2.22   2.42   2.04
   N  .    500    500    500    500    500    500    500    500    500
corr V1  1.000      .      .      .      .      .      .      .      .
corr V2  0.290  1.000      .      .      .      .      .      .      .
corr V3  0.372  0.478  1.000      .      .      .      .      .      .
corr V4  0.587  0.238  0.209  1.000      .      .      .      .      .
corr V5  0.201  0.586  0.126  0.213  1.000      .      .      .      .
corr V6  0.218  0.281  0.681  0.195  0.096  1.000      .      .      .
corr V7  0.557  0.228  0.195  0.664  0.242  0.232  1.000      .      .
corr V8  0.196  0.644  0.146  0.261  0.641  0.248  0.383  1.000      .
corr V9  0.219  0.241  0.676  0.290  0.168  0.749  0.361  0.342  1.000
;
run;

proc calis data= CLUSTA cov method=ml pall pcoves;
var = V1-V9;
lineqs
 V1 = lam1 f1 + e1,
 V2 = lam2 f2 + e2,
 V3 = lam3 f3 + e3,
 V4 = lam4 f1 + e4,
 V5 = lam5 f2 + e5,
 V6 = lam6 f3 + e6,
 V7 = lam7 f1 + e7,
 V8 = lam8 f2 + e8,
 V9 = lam9 f3 + e9;
```

(cont.)

TABLE 6.3. *(cont.)*

```
std
  f1-f3 = 1.0,
  e1-e9 = td1-td9;
cov
  f1-f3 = ph1-ph3,
  e1-e3 = td10-td12,
  e4-e6 = td13-td15,
  e7-e9 = td16-td18;
run;
```

Amos Basic 5.0

```
' Example of Correlated Uniqueness MTMM in Amos 5.0

Sub Main ()
Dim sem As New AmosEngine

sem.TextOutput
sem.Standardized
sem.Mods 4              ' same as Amos default, MIs  4.0 are printed
sem.Smc

sem.BeginGroup "mtmm.txt"

sem.Structure "x1 =   PARANOID + (1) E1"
sem.Structure "x2 =   SCHIZOTP + (1) E2"
sem.Structure "x3 =   SCHIZOID + (1) E3"
sem.Structure "x4 =   PARANOID + (1) E4"
sem.Structure "x5 =   SCHIZOTP + (1) E5"
sem.Structure "x6 =   SCHIZOID + (1) E6"
sem.Structure "x7 =   PARANOID + (1) E7"
sem.Structure "x8 =   SCHIZOTP + (1) E8"
sem.Structure "x9 =   SCHIZOID + (1) E9"

sem.Structure "PARANOID (1)"
sem.Structure "SCHIZOTP (1)"
sem.Structure "SCHIZOID (1)"

sem.Structure "PARANOID  SCHIZOTP"
sem.Structure "PARANOID  SCHIZOID"
sem.Structure "SCHIZOTP  SCHIZOID"

sem.Structure "E1  E2"
sem.Structure "E1 <--> E3"
sem.Structure "E2 <--> E3"
sem.Structure "E4 <--> E5"
sem.Structure "E4 <--> E6"
sem.Structure "E5 <--> E6"
sem.Structure "E7 <--> E8"
sem.Structure "E7 <--> E9"
sem.Structure "E8 <--> E9"

End Sub
```

TABLE 6.3. Computer Syntax (LISREL, Mplus, EQS, CALIS, Amos) for Specification of a Correlated Uniquenesses CFA of the MTMM Matrix of Cluster A Personality Disorders

LISREL 8.72

```
TITLE LISREL SYNTAX FOR CORRELATED UNIQUENESS MTMM SPECIFICATION
DA NI=9 NO=500 MA=CM
LA
PARI SZTI SZDI PARC SZTC SZDC PARO SZTO SZDO
KM
1.000
0.290  1.000
0.372  0.478  1.000
0.587  0.238  0.209  1.000
0.201  0.586  0.126  0.213  1.000
0.218  0.281  0.681  0.195  0.096  1.000
0.557  0.228  0.195  0.664  0.242  0.232  1.000
0.196  0.644  0.146  0.261  0.641  0.248  0.383  1.000
0.219  0.241  0.676  0.290  0.168  0.749  0.361  0.342  1.000
SD
3.61 3.66 3.59 2.94 3.03 2.85 2.22 2.42 2.04
MO NX=9 NK=3 PH=SY,FR LX=FU,FR TD=SY,FR
LK
PARANOID SCHIZOTYP SCHIZOID
PA LX
1 0 0
0 1 0
0 0 1
1 0 0
0 1 0
0 0 1
1 0 0
0 1 0
0 0 1
PA TD
1
1 1
1 1 1
0 0 0 1
0 0 0 1 1
0 0 0 1 1 1
0 0 0 0 0 0 1
0 0 0 0 0 0 1 1
0 0 0 0 0 0 1 1 1
PA PH
0
1 0
1 1 0
VA 1.0 PH(1,1) PH(2,2) PH(3,3)
OU ME=ML RS MI SC AD=OFF IT=500 ND=4
```

(cont.)

TABLE 6.3. *(cont.)*

Mplus 3.11

```
TITLE:     MPLUS PROGRAM FOR CORRELATED UNIQUENESS MTMM MODEL
DATA:      FILE IS "C:\INPUT5.DAT";
           TYPE IS STD CORR;
           NOBS ARE 500;
VARIABLE:  NAMES ARE PARI SZTI SZDI PARC SZTC SZDC PARO SZTO SZDO;
ANALYSIS:  ESTIMATOR=ML;
MODEL:     PARANOID BY PARI* PARC PARO;
           SCHZOTYP BY SZTI* SZTC SZTO;
           SCHIZOID BY SZDI* SZDC SZDO;
           PARANOID@1.0; SCHZOTYP@1.0; SCHIZOID@1.0;
           PARI SZTI WITH SZDI; PARI WITH SZTI; PARC SZTC WITH SZDC;
           PARC WITH SZTC; PARO SZTO WITH SZDO; PARO WITH SZTO;
           PARANOID SCHZOTYP WITH SCHIZOID; PARANOID WITH SCHZOTYP;
OUTPUT:    SAMPSTAT MODINDICES(3.84) STAND RESIDUAL;
```

EQS 5.7b

```
/TITLE
 EQS SYNTAX FOR CORRELATED UNIQUENESS MTMM SPECIFICATION
/SPECIFICATIONS
 CASES=500; VARIABLES=9; METHODS=ML; MATRIX=COR; ANALYSIS=COV;
/LABELS
 v1=PARI; v2=SZTI; v3=SZDI; v4=PARC;
 v5=SZTC; v6=SZDC; v7=PARO; v8=SZTO; v9=SZDO;
 f1 = paranoid; f2 = schizotyp; f3 = schizoid;
/EQUATIONS
 V1 =  *F1+E1;
 V2 =  *F2+E2;
 V3 =  *F3+E3;
 V4 =  *F1+E4;
 V5 =  *F2+E5;
 V6 =  *F3+E6;
 V7 =  *F1+E7;
 V8 =  *F2+E8;
 V9 =  *F3+E9;
/VARIANCES
 F1 TO F3 = 1.0;
 E1 TO E9= *;
/COVARIANCES
 F1 TO F3 = *;
 E1 TO E3= *; E4 TO E6= *; E7 TO E9= *;
/MATRIX
1.000
0.290  1.000
0.372  0.478  1.000
0.587  0.238  0.209  1.000
0.201  0.586  0.126  0.213  1.000
0.218  0.281  0.681  0.195  0.096  1.000
0.557  0.228  0.195  0.664  0.242  0.232  1.000
0.196  0.644  0.146  0.261  0.641  0.248  0.383  1.000
0.219  0.241  0.676  0.290  0.168  0.749  0.361  0.342  1.000
```

(cont.)

tions), results of the correlated methods CFA specification for the Cluster A personality disorders example are not provided.

An important special case of correlated methods solutions is the *uncorrelated methods* model. Its specification is identical to that of correlated methods, except that the covariances of the methods factors are fixed to zero (Widaman, 1985); that is, in the LISREL syntax provided in Table 6.2, this would simply entail fixing the PH(5,4), PH(6,4), and PH(6,5) elements to zero instead of freely estimating these parameters. Because these two models are nested, this comparison provides a statistical evaluation of whether the effects associated with the different assessment methods are correlated; for example, a lack of correlated method effects would be indicated by a nonsignificant χ^2 difference test.

Correlated Uniqueness Models

The correlated uniqueness CFA model (Kenny, 1979; Marsh, 1989) was introduced as an alternative approach to analyzing MTMM data. Figure 6.2 depicts the path diagram of the correlated uniqueness CFA specification for the MTMM matrix of Cluster A personality disorders (Table 6.1). In order for the correlated uniqueness model to be identified, there must be at least two traits (T) and three methods (M) (although a $2T \times 2M$ model can be fit to the data if the factor loadings of indicators loading on the same trait factor are constrained to equality). Specification of the trait por-

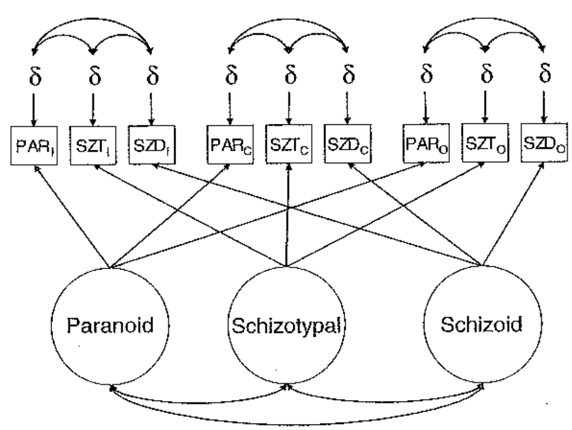


FIGURE 6.2. Correlated uniqueness CFA specification of the MTMM matrix of Cluster A personality disorders.

tion of the correlated uniqueness model is the same as that of the cor lated methods approach: (1) each indicator is specified to load on one tr factor (all other cross-loadings are fixed to zero); and (2) correlatio among trait factors are freely estimated. Thus, the key difference betwe these parameterizations is the manner in which method effects are es mated. In the correlated uniqueness model, method effects are estimat by specifying correlated uniquenesses (errors) among indicators based the same assessment method rather than by method factors. Table 6.3 pr vides the programming syntax for this model in the LISREL, Mplus, EC CALIS, and Amos languages.

A strong practical advantage of the correlated uniqueness approa is that, unlike correlated methods models, this parameterization rare results in improper solutions (Kenny & Kashy, 1992; Marsh & Bail 1991; Tomás, Hontangas, & Oliver, 2000). Indeed, the correlated uniqu ness MTMM model of Cluster A personality disorders converged and pr duced a proper solution that provided an acceptable fit to the data, χ^2 (1 = 14.34, p = .50, SRMR = .025, RMSEA = 0.00 (90% CI = 0.00 to 0.04, Cl = .99), TLI = 1.00, CFI = 1.00. Inspection of standardized residuals a modification indices indicated a reasonable solution. The completely sta dardized parameter estimates of this solution are presented in Table 6 With the exception of one correlated uniqueness (δ_{54}), all unstandardiz parameter estimates are significantly different from zero ($p < .05$).

In regard to the parameter estimates produced by a correlated uniqu ness CFA model specification, trait factor loadings that are large and stat tically significant would be viewed in support of convergent validi However, large correlations among the trait factors would be indicative poor discriminant validity. The presence of appreciable method effects reflected by correlated uniquenesses among indicators assessed by t same method that are moderate or greater in magnitude. The results Table 6.4 support the construct validity of the Cluster A personality dis ders. The trait factor loadings are consistently large (range = .712–.87 providing evidence that the indicators are strongly related to their pu ported latent constructs (convergent validity), adjusting for the effects assessment method. Adequate discriminant validity is evidenced by t modest correlations among trait factors (range = .310–.381). Although si nificant method effects were obtained in all but one instance (SZT_C wi PAR_C), the size of these effects is modest (standardized values range –.037–.293). As in the MTMM matrix presented in Table 6.1, meth effects estimated by the CFA model are smallest for the clinical ratin indicators.

Correlated Methods Models

The correlated methods parameterization reflects the traditional CFA approach to analyzing MTMM matrices. There are five major aspects of correlated methods model specifications (Widaman, 1985): (1) to be identified, there must be at least three traits (T) and three methods (M); (2) $T \times M$ indicators are used to define $T + M$ latent factors (i.e., the number of trait factors = T; the number of method factors = M); (3) each indicator is specified to load on two factors—its trait factor and its method factor (all other cross-loadings are fixed to zero); (4) correlations among trait factors and among method factors are freely estimated, but the correlations between trait and method factors are usually fixed to zero; and (5) indicator uniquenesses (i.e., variance in the indicators not explained by the trait and method factors) are freely estimated but cannot be correlated with the uniquenesses of other indicators. Accordingly, in this specification, each indicator is considered to be a function of trait, method, and unique factors.

Figure 6.1 depicts the path diagram of the correlated methods CFA specification for the MTMM matrix of Cluster A personality disorders (Table 6.1). Table 6.2 provides the LISREL syntax for this model. For reasons discussed later in this chapter (e.g., propensity for improper solu-

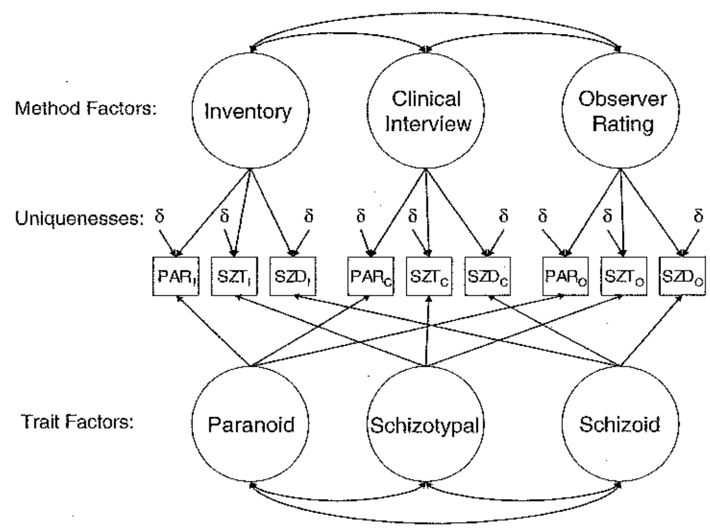


FIGURE 6.1. Correlated methods CFA specification of the MTMM matrix of Cluster A personality disorders.

TABLE 6.2. LISREL Syntax for Specification of a Correlated Methods CFA of the MTMM Matrix of Cluster A Personality Disorders

```
TITLE LISREL SYNTAX FOR CORRELATED METHODS MTMM SPECIFICATION
DA NI=9 NO=500 MA=CM
LA
PARI SZTI SZDI PARC SZTC SZDC PARO SZTO SZDO
KM
1.000
0.290  1.000
0.372  0.478  1.000
0.587  0.238  0.209  1.000
0.201  0.586  0.126  0.213  1.000
0.218  0.281  0.681  0.195  0.096  1.000
0.557  0.228  0.195  0.664  0.242  0.232  1.000
0.196  0.644  0.146  0.261  0.641  0.248  0.383  1.000
0.219  0.241  0.676  0.290  0.168  0.749  0.361  0.342  1.000
SD
3.61 3.66 3.59 2.94 3.03 2.85 2.22 2.42 2.04
MO NX=9 NK=6 PH=SY,FR LX=FU,FR TD=SY,FR
LK
PARANOID SCHIZOTYP SCHIZOID INVENTRY INTERVW OBSERVE
PA LX
1 0 0 1 0 0
0 1 0 1 0 0
0 0 1 1 0 0
1 0 0 0 1 0
0 1 0 0 1 0
0 0 1 0 1 0
1 0 0 0 0 1
0 1 0 0 0 1
0 0 1 0 0 1
PA TD
1
0 1
0 0 1
0 0 0 1
0 0 0 0 1
0 0 0 0 0 1
0 0 0 0 0 0 1
0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1
PA PH
0
1 0
1 1 0
0 0 0 0
0 0 0 1 0
0 0 0 1 1 0
VA 1.0 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5) PH(6,6)
OU ME=ML RS MI SC AD=OFF IT=500 ND=4
```

tional ratings made by paraprofessional staff. Thus, Table 6.1 is a 3 (*T*) × 3 (*M*) matrix, arranged such that the correlations among the different traits (personality disorders: paranoid, schizotypal, schizoid) are nested within each method (assessment type: inventory, clinical interview, observer ratings). The MTMM is a symmetric correlation matrix with one exception: reliability estimates (e.g., Cronbach's alphas) of the measures are inserted in the diagonal in place of ones (e.g., in Table 6.1, the internal consistency estimate of the inventory measure of paranoid personality is .93). As Campbell and Fiske (1959) note, ideally the reliability diagonal should contain the largest coefficients in the matrix; that is, the measure should be more strongly correlated with itself than with any other indicator in the MTMM matrix.

The MTMM matrix consists of two general types of blocks of coefficients: (1) *monomethod blocks*, which contain correlations among indicators derived from the same assessment method; and (2) *heteromethod blocks*, which contain correlations among indicators assessed by different methods (see Table 6.1). Of central interest is the *validity diagonal*, which corresponds to the diagonal within each heteromethod block—the bolded correlations in Table 6.1. Correlations on the validity diagonal represent estimates of *convergent validity*: different measures of theoretically similar or overlapping constructs should be strongly interrelated. In the MTMM matrix, convergent validity is evidenced by strong correlations among methods measuring the same trait (i.e., monotrait–heteromethod coefficients). For example, the findings in Table 6.1 indicate that the three different measures of schizotypal personality are strongly interrelated (range of *r*s = .676 to .749). The off-diagonal elements of the heteromethod blocks reveal *discriminant validity*: measures of theoretically distinct constructs should not be highly intercorrelated. Discriminant validity in the MTMM matrix is evidenced by weaker correlations between different traits measured by different methods (i.e., heterotrait–heteromethod coefficients) in relation to correlations on the validity diagonal (monotrait–heteromethod coefficients). In Table 6.1, support for discriminant validity is obtained by the finding that correlations in the off-diagonal elements of the heteromethod blocks are uniformly lower (range of *r*s = .126 to .290) than the validity coefficients (range of *r*s = .557 to .749).

Finally, evidence of *method effects* is obtained by an examination of the off-diagonal elements of the monomethod blocks. The extent of method effects is reflected by the differential magnitude of correlations between

different traits measured by the same method (heterotrait–monometho coefficients) relative to the correlations between the same two traits mea sured by different methods. As shown in Table 6.1, although not extrem some method variance is evident, especially for the inventory and observe rating measures. For example, the observer ratings of the traits of paranoi and schizotypal personality are more highly correlated (r = .383) tha heteromethod measures of these traits (e.g., the correlation betwee paranoid and schizotypal personality traits measured by inventory an observer rating, respectively, is .196; see Table 6.1). As in the prese example, when the collective results indicate that convergent and discrim nant validity are high and method effects are negligible, construct validi is supported.

CFA APPROACHES TO ANALYZING THE MTMM MATRIX

Despite the fact that the Campbell and Fiske (1959) methodology repre sented a significant advance in the conceptualization and evaluation c construct validity, the MTMM approach was not widely used over th years immediately following its inception. In addition, several limitation of the MTMM methodology were noted, including the subjective nature c its interpretation (e.g., ambiguity in terms of what patterns of correlation reflect satisfactory convergent and discriminant validity), its reliance o correlations among fallible observed measures to draw inferences abou trait and methods factors (cf. Schmitt & Stults, 1986), and the failure c EFA to obtain meaningful solutions of MTMM data (e.g., identificatio restrictions of EFA prevent specification of correlated errors; see Cha ter 2).

Interest in the MTMM matrix increased with the realization that th procedures of CFA could be readily applied to its analysis (cf. Cole, 198 Flamer, 1983, Marsh & Hocevar, 1983; Widaman, 1985). In other word MTMM matrices, like any other form of correlation or covariance matri can be analyzed by CFA to make inferences about potential underlyin dimensions such as trait and methods factors. Although several differe types of CFA models can be applied to MTMM data (cf. Marsh & Grayso 1995; Widaman, 1985), two forms of CFA specification have been pr dominant. Using more contemporary terminology, these two types of sol tions have been referred to as *correlated methods* and *correlated uniquene* models (Marsh & Grayson, 1995).

negative symptom rating scale could be attributed to artifacts of the indicator set. For example, the chances of obtaining a distinct latent factor of Flat Affect may be fostered by creation of a rating scale containing several similarly worded items that assess this feature. This issue is particularly salient in instances where factor analysis uncovers more latent dimensions than initially predicted; that is, are these additional dimensions conceptually and practically useful, or do they stem from artifacts of scale development (cf. Models A and B in Figure 5.3 in Chapter 5)? In addition, method effects may obscure the discriminant validity of the constructs. That is, when each construct is assessed by the same measurement approach (e.g., observer rating), it cannot be determined how much of the observed overlap (i.e., factor correlations) is due to method effects as opposed to "true" covariance of the traits. In sum, construct validation is limited in instances where a single assessment method is employed.

Campbell and Fiske (1959) developed the multitrait–multimethod (MTMM) matrix as a method for establishing the construct validity of psychological measures. This methodology entails a matrix of correlations arranged in a manner that fosters the evaluation of construct validity. *Construct validity* is the overarching principle of validity, referring to the extent to which a psychological measure in fact measures the concept it purports to measure. This approach requires that several *traits* (*T*; e.g., attitudes, personality characteristics, behaviors) are each assessed by several *methods* (*M*; e.g., alternative test forms, alternative assessment modalities such as questionnaires and observer ratings, or separate testing occasions). The result is a $T \times M$ correlation matrix that is interpreted with respect to convergent validity, discriminant validity, and method effects.

An example of the MTMM matrix is presented in Table 6.1. In this illustration, the researcher wishes to examine the construct validity of the DSM-IV Cluster A personality disorders, which are enduring patterns of symptoms characterized by odd or eccentric behaviors (American Psychiatric Association, 1994). Cluster A is comprised of three personality disorder constructs: (1) paranoid (an enduring pattern of distrust and suspicion such that others' motives are interpreted as malevolent); (2) schizoid (an enduring pattern of detachment from social relationships and restricted range of emotional expression); and (3) schizotypal (an enduring pattern of acute discomfort in social relationships, cognitive and perceptual distortions, and behavioral eccentricities). In a sample of 500 patients, each of these three traits is measured by three assessment methods: (1) a self-report inventory of personality disorders; (2) dimensional ratings from a structured clinical interview of personality disorders; and (3) observa-

TABLE 6.1. Multitrait–Multimethod Matrix of Cluster A Personality Disorder Constructs

	Inventory			Clinical Interview			Observer Ratings		
	PAR	SZT	SZD	PAR	SZT	SZD	PAR	SZT	SZD
Inventory									
Paranoid (PAR)	(.930)								
Schizotypal (SZT)	.290	(.900)							
Schizoid (SZD)	.372	.478	(.910)						
Clinical Interview									
Paranoid (PAR)	**.587**	.238	.209	(.800)					
Schizotypal (SZT)	.201	**.586**	.126	.213	(.870)				
Schizoid (SZD)	.218	.281	**.681**	.195	.096	(.830)			
Observer Ratings									
Paranoid (PAR)	**.557**	.228	.195	**.664**	.242	.232	(.820)		
Schizotypal (SZT)	.196	**.644**	.146	.261	**.641**	.248	.383	(.790)	
Schizoid (SZD)	.219	.241	**.676**	.290	.168	**.749**	.361	.342	(.800)
Standard Deviations:	3.610	3.660	3.590	2.940	3.030	2.850	2.220	2.420	2.040

Note. Values on the diagonal in parentheses are indicator reliabilities; correlations in boldface type represent convergent validities; single-line boxes are heteromethod blocks; double-line boxes are monomethod blocks; correlations in monomethod blocks represent heterotrait–monomethod coefficients; nonbolded correlations in heteromethod blocks represent heterotrait–heteromethod coefficients.

6

CFA of Multitrait–Multimethod Matrices

As seen in prior chapters, CFA provides an elegant analytic framework for evaluating the validity of constructs, and for examining the interrelations of constructs adjusting for measurement error and an error theory. The application of CFA to multitrait–multimethod (MTMM) matrices offers an even more sophisticated methodology for estimating convergent validity, discriminant validity, and method effects in the evaluation of the construct validity of constructs in the social and behavioral sciences. This chapter discusses the various ways an MTMM CFA model can be parameterized and extended, and the strengths and drawbacks of each approach. In addition to bolstering the concepts of convergent and discriminant validity in the context of CFA, the chapter illustrates the deleterious consequences that may result from failing to account for measurement error and method effects.

CORRELATED VERSUS RANDOM MEASUREMENT ERROR REVISITED

In Chapter 5, the need for specifying correlated errors was discussed in instances where some of the observed covariance of two or more indicators is believed to be the effect of the measurement approach (i.e., method covariance), over and above covariance explained by the substantive latent factors. Thus, the error theory of a CFA measurement model can entail some combination of *random measurement error* (i.e., the unexplained vari-

ance of one indicator does not covary with the unexplained variance of another) and *correlated measurement error* or uniqueness (i.e., the unexplained variance of one indicator covaries with the unexplained variance of another), as guided by the conceptual or empirical basis of the model specification. As discussed in Chapter 5, correlated errors should not be specified solely for the purpose of improving model fit. In the majority of models involving latent variables defined by multiple indicators, measurement errors are freely estimated (and correlated errors, if applicable). However, in some instances it is justified to impose constraints on these parameters (e.g., constraining error variances to equality, as in the evaluation of parallel tests) or to fix these estimates to predetermined values (e.g., prespecifying the amount of measurement error in a variable measured by a single indicator; Chapter 4). These alternative specifications for measurement errors are discussed in other chapters (Chapter 4 and Chapter 7).

THE MULTITRAIT–MULTIMETHOD MATRIX

A common limitation of applied research is that the dimensionality and validity of constructs are evaluated in a cross-sectional fashion using a single measurement scale. For example, a researcher may hypothesize that the negative symptoms of schizophrenia are comprised of three components: flat affect, alogia (poverty of speech), social amotivation. To examine this notion, he or she develops a multi-item clinical observation rating system that is subsequently used to rate the behavior of patients admitted to a state hospital. After a sufficient sample has been collected, the ratings are submitted to factor analysis. The results indicate a three-factor solution, which the researcher interprets as supporting the conjectured tripartite model of negative symptoms. He or she concludes that the findings attest to favorable convergent and discriminant validity in that features of flat affect, alogia, and social amotivation are differentially intercorrelated to such an extent that they load on distinct latent factors. Further support for validity is obtained by results showing that the three latent factors are more strongly related to measures of schizophrenia severity than indicators of other disorders (e.g., bipolar disorder).

The aforementioned scenario is a common empirical sequence in scale development and construct validation. Although a useful part of such endeavors, this sequence provides an incomplete evaluation of construct validity. It is not clear as to what extent the multidimensionality of the respecified model) and methodological considerations pertaining to the principles of model identification, equivalent models, model comparison, and improper solutions (e.g., knowing when additional parameters will lead to underidentification).

The first five chapters of this book have covered what could be considered the "fundamentals" of CFA; that is, assuming no complications such as non-normality or missing data, what every researcher "needs to know" to conduct CFA properly with applied data sets. Beginning with Chapter 6, which focuses on the topic of CFA with multitrait–multimethod matrices, the remaining chapters of this book address more specialized applications of CFA (e.g., higher-order factor models, multiple-groups solutions, MIMIC models, scale reliability evaluation) and other issues that frequently arise in the analysis of real data (e.g., missing data, non-normality, categorical indicators, statistical power analysis).

NOTES

1. Correlated errors are sometimes referred to as *minor factors*, especially in instances where three or more indicators loading a broader latent factor have intercorrelated measurement errors. This terminology is most appropriate when the error covariances are presumed to reflect a substantively salient dimension (rather than a byproduct of the measurement approach such as reverse-worded items) that is subsumed by a broader latent construct.

2. Unlike factor correlations = 1.0 or error variances = 0.0 (see discussion of comparing one- and two-factor models), factor loadings, factor covariances, and indicator error covariances that are fixed to 0.0 are not restrictions on the edge of inadmissible parameter spaces (i.e., theoretically, these unstandardized parameters can have values of $\pm \infty$).

3. Interested readers are referred to Yuan, Marshall, and Bentler (2003) for a detailed demonstration of how model misspecification affects parameter biases.

4. An alternative, yet less frequently used, approach to comparing non-nested models involves tests based on "nested tetrads" (Bollen & Ting, 1993, 2000). Tetrads are differences in the products of pairs of covariances (e.g., $\tau_{1234} = \sigma_{12}\sigma_{34} - \sigma_{13}\sigma_{24}$). Depending on the model specification, some tetrads will equal zero (referred to as "vanishing tetrads"). Bollen (1990) derived a χ^2 statistic to perform a simultaneous test of vanishing tetrads of a model to assess its fit to data (*df* = the number of vanishing tetrads). Bollen and Ting have demonstrated scenarios where two models are not structurally nested (i.e., one model does not contain a subset of the freed parameters of the other) but are nested in their vanishing tetrads (i.e., the vanishing tetrads of one model are a subset of the other's). In these instances, two models can be compared by taking the difference in their χ^2s.

5. This rule also applies to weight matrices associated with non-normal theory estimators (e.g., weighted least squares, Chapter 9) and variance–covariance

6. On a related note, Heywood cases may occur in EFA when too many factors have been extracted from the data.

7. However, constraining a variance to zero places the estimate on the border of inadmissibility (see the discussion in this chapter on the use of χ^2 difference testing to compare CFA models that differ in number of factors).

8. Although the most widely used software programs for EFA (e.g., SPSS) do not provide significance testing of factor loadings, this feature is available in the Comprehensive Exploratory Factor Analysis freeware program developed by Michael Browne (this can be downloaded at quantrm2.psy.ohio-state.edu/browne/software.htm), and in LISREL 8.72.

9. On rare occasions, equal goodness of fit (in terms of χ^2, CFI, etc.) may occur by chance for two models that do not produce the sample model-implied covariance matrix. This is not an instance of equivalent models.

10. As shown in Chapter 8, a hierarchical model involving a single higher-order factor and three lower-order factors produces the same goodness of fit as a three-factor model where the factors are freely intercorrelated. This is because the higher-order portion of the hierarchical model is just-identified (i.e., analogous to Model B in Figure 3.6, Chapter 3).

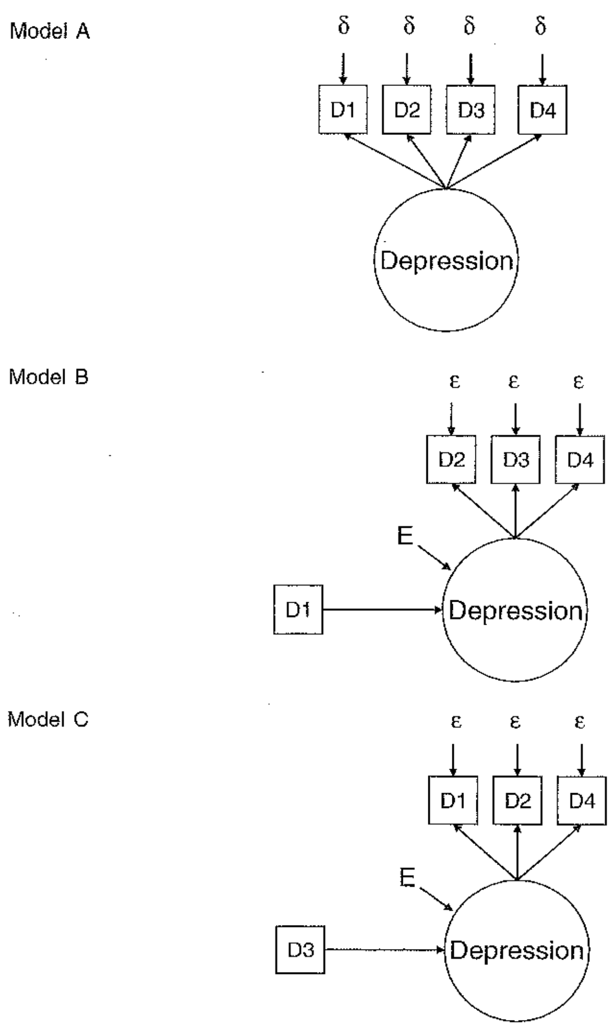


FIGURE 5.4. Examples of equivalent CFA models of depression. D1, hopelessness; D2, depressed mood; D3, guilt; D4, loss of interest in activities.

Social Psychology between 1988 and 1991, 100% reported a model wher there existed three or more equivalent models. The median number c equivalent models in these studies was 12 (range of equivalent models = to 33,925!). However, none of these studies contained an acknowledgmer of the existence of equivalent solutions. In the 10+ years since the publica tion of MacCallum et al. (1993), little has changed in regard to applie researchers' explicit recognition of equivalent models. This may be due i part to the unavailability of utilities in latent variable software packages t generate the equivalent models associated with the models specified by th researcher. If such utilities were available, the equivalent models reveale by the software could be evaluated by the researcher in terms of their sub stantive plausibility. In some cases, this process may lend further suppor for the hypothesized model (i.e., in situations where there exist no zer equivalent models, or where each of the equivalent models is conceptuall not viable). In other instances, the process may have considerable heuris tic value (e.g., reveal theoretically plausible alternative models that wer not recognized by the researcher).

SUMMARY

Poor-fitting models are common in applied data sets. It is important t have a sound knowledge of the sample data and model before proceedin with a CFA. Data screening procedures (e.g., normality, outliers), in tan dem with a proper missing data strategy (see Chapter 9), should be con ducted to help ensure that the sample data are appropriate for CFA and th statistical estimator (e.g., ML). Principal components analysis is anothe helpful procedure in this phase to verify that the input matrix is positiv definite (i.e., all resulting eigenvalues are positive). EFA and E/CFA ar important precursors to CFA for developing and refining a measuremen model. Although use of these methods will foster the chances of success i the restrictive CFA framework (e.g., most or all cross-loadings and erro covariances are fixed to zero), the solution still may not satisfy one o more of the three major criteria for model acceptability (i.e., overall good ness of fit, areas of localized strain in the solution, interpretability/strengt or parameter estimates). Thus, the researcher must be adept at diagnosin and rectifying the sources of ill fit (interpretation of modification indices standardized residuals, and reasonability of parameter estimates), keepin in mind both substantive issues (i.e., conceptual justification for th

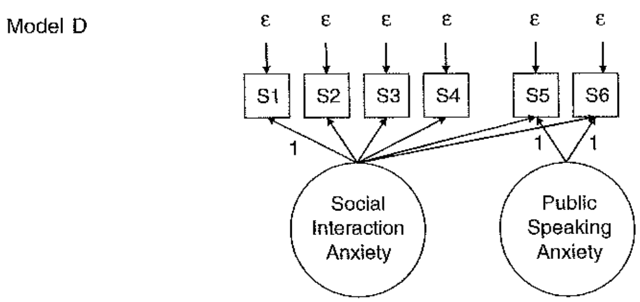


FIGURE 5.3. *(cont.)*

Figure 5.3 presents two additional equivalent CFA models of social anxiety (assuming that a two-factor solution is conceptually viable). Model C depicts a higher-order factor model (see Chapter 8) in which the correlation between the lower-order factors of Social Interaction Anxiety and Public Speaking Anxiety (Model A) is accounted for by a second-order factor of General Social Anxiety. This specification would imply that social interaction anxiety and public speaking anxiety represent distinct subdimensions influenced by the broader construct of general social anxiety. It should be noted that because there are only two lower-order factors, the higher-order portion of Model C would be underidentified if the loadings of General Anxiety → Social Interaction Anxiety and General Social Anxiety → Public Speaking Anxiety were freely estimated (i.e., it is analogous to the underidentified model depicted by Model A in Figure 3.6, Chapter 3). Thus, to identify Model C, the two higher-order factor loadings must be constrained to be equal (equality constraints are discussed in detail in Chapter 7).[10]

Methodologists have noted that some equivalent solutions can be readily dismissed based on logic or theory (e.g., MacCallum, Wegener, Uchino, & Fabrigar, 1993). Model D could be regarded as one such example. In this model, S5 and S6 are purported to be indicators of both Social Interaction Anxiety and Public Speaking Anxiety (this specification requires an equality constraint for the S5 and S6 loadings on Public Speaking Anxiety). However, Social Interaction Anxiety and Public Speaking Anxiety are assumed to be unrelated (i.e., the factor covariance is fixed to zero). It is likely that both of these assertions (S5 and S6 as indicators of anxiety in one-on-one social interactions; orthogonal nature of social interaction and public speaking anxiety) can be quickly rejected on conceptual grounds. Other examples of nonsensical equivalent solutions include longitudinal models with directional relationships among temporal variables (e.g., regressing a Time 1 variable onto a Time 2 variable; i.e., Time 1 → Time 2 vs. Time 2 → Time 1) or models with direct effects of predetermined exogenous indicators such as gender or age (e.g., Gender → Job Satisfaction vs. Job Satisfaction → Gender).

Figure 5.4 presents three equivalent solutions that entail a single latent factor. In this example, the four indicators are observed measures related to the construct of depression: hopelessness (D1), depressed mood (D2), feelings of guilt (D3), and loss of interest in usual activities (D4). In Model A, a simple one-factor measurement model is specified in which the four indicators are presumed to be interrelated because they are each influenced (caused) by the underlying dimension of Depression. In Model B, the latent factor of Depression (now defined by indicators D2–D4) is regressed onto the single indicator of hopelessness (D1). Although providing the same fit to the data, Models A and B are profoundly different in terms of their conceptual implications. Unlike Model A, which presumes that hopelessness is just another manifest symptom of depression, Model B is in accord with a conceptualization that a sense of hopelessness is a cause of depression. Note that Model A could be respecified such that any of the four indicators are designated as having direct effects on the Depression latent variable (e.g., Model C is similar to Model B, except that guilt, D3, is specified as the cause of Depression). Models B and C can be regarded as examples of CFA with covariates (covariates are sometimes referred to as background variables; cf. Chapter 7). In such models, the latent factors are endogenous (i.e., Latent Y variables) because the solution attempts to explain variance in them with an exogenous indicator (e.g., D1 in Model B). Accordingly, in Models B and C, the residual variance in Depression (often referred to as a *disturbance*; see Chapters 3 and 7) is freely estimated in the psi matrix (cf. Model A, where the factor variance is estimated in the phi matrix). The residual variance in Depression is depicted as "E" in the path diagrams of Models B and C (i.e., "E" = φ_1). As will be seen in Chapter 9, the issue of equivalent models is also germane to solutions that contain formative constructs (i.e., latent variables "caused" by a composite set of indicators), because such models can often be alternatively parameterized as MIMIC models (Chapter 7).

Perhaps the most widely cited article on the issue of equivalent SEM models was published by MacCallum and colleagues in 1993. These authors found that although equivalent models were quite common in applied SEM research, this issue was virtually ignored by investigators. For example, of the 20 SEM articles published in the *Journal of Personality and* greater number of alternative, equivalent solutions than more parsimonious models.

Figure 5.3 presents four equivalent CFA solutions. In this example, the researcher is interested in examining the latent dimensionality of situational social anxiety (i.e., anxiety in social situations due to concerns about being negatively evaluated by others, embarrassment, etc.). Of the six indicators, four could be construed as measuring anxiety in one-on-one social interactions (S1–S4), and the remaining two pertain to anxiety in public speaking situations (S5, S6). Each of the four CFA models in Figure 5.3 is overidentified, with $df = 8$, and would fit the data equally well.

Models A and B exemplify a quandary often faced by applied CFA researchers—in this case, should the additional covariance existing between the two public speaking indicators be accounted for as a distinct latent factor (Model A), or within a unidimensional solution containing an error covariance between S5 and S6 (Model B)? These alternative specifications may have substantial conceptual implications in the researcher's applied area of study. For instance, the Model A specification would forward the conceptual notion that the construct of Social Anxiety is multidimensional. Model B would assert that Social Anxiety is a broader unitary construct and that the differential covariance of indicators S5 and S6 represents a correlated error, perhaps due to a substantively trivial method effect (unlike S1–S4, these two indicators are more specifically worded to assess anxiety in large groups). Because Models A and B provide the same fit to the data, the procedures of CFA cannot be employed to resolve the question of which model is more acceptable. This is particularly problematic in instances where two or more equivalent models are substantively plausible.

This example also illustrates how model fit and equivalent solutions can be impacted by the composition of the indicator set. The latent dimensionality of a collection of indicators may be strongly impacted by potentially artifactual issues, such as the inclusion of individual items with very similar or reverse wordings or the use of highly overlapping (multicollinear) measures. In the present case, it might be decided prior to specifying a unidimensional CFA model of social anxiety (along the lines of Model B) that the S5 and S6 indicators are over-redundant (i.e., both assess anxiety in speaking in front of large groups). If one of these indicators was omitted from the analyses (or if the S5 and S6 were combined; cf. parceling in Chapter 9), the researcher would avert the problems of poor model fit (if a correlated error had not been specified between S5 and S6 in an initial one-factor

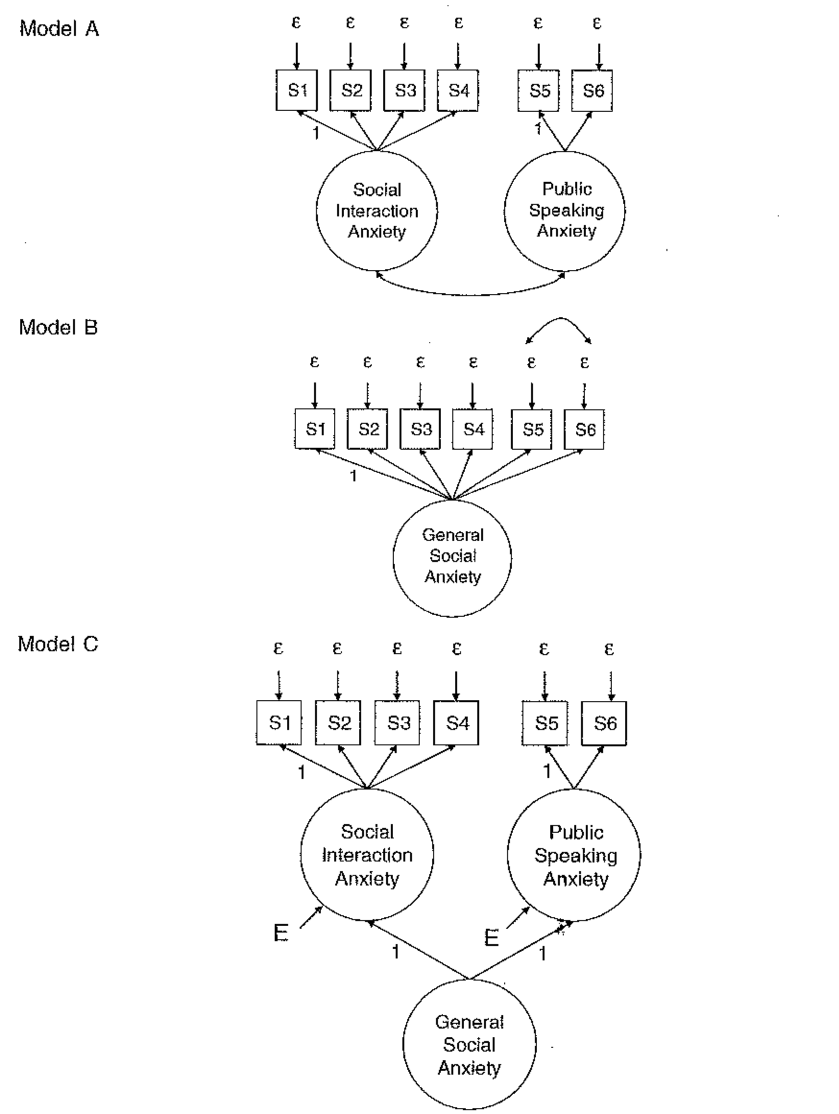


FIGURE 5.3. Examples of equivalent CFA models of social anxiety. S1, makii eye contact; S2, maintaining a conversation; S3, meeting strangers; S4, speakii on the telephone; S5, giving a speech to a large group; S6, introducing yourself large groups. All path diagrams use latent-Y notation. In order for Model C to l identified, the two higher-order factor loadings (i.e., General Social Anxiety - Social Interaction Anxiety, General Social Anxiety → Public Speaking Anxiet

As shown in Table 5.8, large modification indices were obtained in regard to the correlated errors of items 11 and 12, and items 9 and 10 (completely standardized EPCs = .28 and .20, respectively). The correlated error between items 11 and 12 may be grounded substantively by the fact that these are the only reverse-worded items in the questionnaire. It was also described earlier in this chapter why the failure to specify this correlated error may negatively impact the ability of the solution to reproduce the observed relationship between items 9 and 10. Accordingly, the researcher would likely take these results as evidence for the need to freely estimate an error covariance between items 11 and 12 in a subsequent CFA solution. The absence of other large modification indices suggests that other measurement errors can be presumed to be random. Collectively, these findings may foster the refinement of the solution initially suggested by EFA—namely, that the CFA model should not be specified as congeneric (i.e., item 4 should load on two factors) and that all error covariances may be fixed to zero with the exception of items 11 and 12.

MODEL IDENTIFICATION REVISITED

General rules and guidelines for CFA model identification were discussed in Chapter 3. That discussion included basic principles such as the need to define a metric for the latent variables, the need for model *df* to be positive (i.e., the number of elements of the input matrix should equal or exceed the number of freely estimated parameters), and the issue of empirical underidentification (e.g., the marker indicator must be significantly related to its latent factor). However, in the current chapter, more complex measurement models have been considered that entail double-loading indicators and correlated indicator errors. Although these general rules still apply, researchers are more likely to encounter identification problems with more complex solutions. Thus, identification issues with such models are now discussed briefly.

Because latent variable software programs are capable of evaluating whether a given model is identified, it is often most practical to simply try to estimate the solution and let the computer determine the model's identification status. Nevertheless, it is helpful for the researcher to be aware of general model identification guidelines to avoid pursuing structurally or empirically underidentified solutions (such as proposing them in research grant applications). In addition to the guidelines presented in Chapter 3, the researcher should be mindful of th

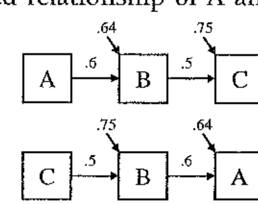
when model *df* is positive). As a general rule, for every indicator there should be at least one other indicator in the solution (which may or may not load on the same latent factor) with which it does not share an error covariance (although in some cases, models that satisfy this rule may still be underidentified; cf. Kenny, Kashy, & Bolger, 1998). In addition, models specified with double-loading indicators may be more prone to underidentification. As discussed in Chapter 6, empirical underidentification is a serious problem in *correlated methods* approaches to CFA multitrait–multimethod analyses where each indicator loads on both a trait factor and a methods factor. The risk for underidentification is increased in models that contain some mixture of double-loading indicators *and* correlated errors. For instance, a solution would be underidentified if an indicator, X1, was specified to load on Factors A and B, but was also specified to have correlated errors with each of the Factor B indicators.

EQUIVALENT CFA SOLUTIONS

Another important consideration in model specification and evaluation is the issue of *equivalent solutions*. Equivalent solutions exist when different model specifications produce identical goodness of fit (with the same number of *df*) and predicted covariance matrices (Σ) in any given data set. Consider the two mediation models depicted below. Both models are overidentified with one *df* corresponding to the nontautological relationship between A and C. Although these models differ greatly in terms of their substantive meaning (Model 1: A is a direct cause of B and an indirect cause of C; Model 2: C is a direct cause of B and an indirect cause of A), they generate the same predicted covariance matrix; that is, in both solutions, the model-implied relationship of A and C is .30.[9]

Model 1:

Model 2:



Interested readers are referred to Stelzl (1986), Lee and Hershberger (1990), and Hershberger (1994), who have developed rules for explicating equivalent solutions for various types of SEM models. A noteworthy aspect

TABLE 5.8. Selected Mplus Output of the E/CFA of the Drinking Motives Questionnaire Items

```
TESTS OF MODEL FIT

Chi-Square Test of Model Fit

          Value                              55.546
          Degrees of Freedom                     33
          P-Value                            0.0083

CFI/TLI

          CFI                                 0.986
          TLI                                 0.972

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                            0.037
          90 Percent C.I.                     0.019  0.053
          Probability RMSEA <=.05             0.898

SRMR (Standardized Root Mean Square Residual)

          Value                               0.023

MODEL RESULTS

                     Estimates     S.E.  Est./S.E.     Std     StdYX

 COPING   BY
    X1                 0.613      0.060    10.274    0.613     0.567
    X2                 0.514      0.075     6.888    0.514     0.497
    X3                 0.487      0.074     6.598    0.487     0.471
    X4                 0.536      0.068     7.873    0.536     0.515
    X5                 0.097      0.068     1.427    0.097     0.100
    X6                 0.114      0.067     1.691    0.114     0.117
    X7                 0.110      0.071     1.556    0.110     0.109
    X8                 0.000      0.000     0.000    0.000     0.000
    X9                 0.032      0.075     0.431    0.032     0.031
    X10               -0.008      0.076    -0.112   -0.008    -0.008
    X11                0.150      0.076     1.957    0.150     0.146
    X12                0.000      0.000     0.000    0.000     0.000

 SOCIAL   BY
    X1                 0.000      0.000     0.000    0.000     0.000
    X2                 0.101      0.076     1.338    0.101     0.098
    X3                 0.105      0.074     1.413    0.105     0.102
    X4                 0.546      0.067     8.209    0.546     0.525
    X5                 0.574      0.055    10.441    0.574     0.594
    X6                 0.651      0.053    12.252    0.651     0.672
    X7                 0.649      0.056    11.492    0.649     0.645
    X8                 0.791      0.046    17.225    0.791     0.753
    X9                 0.020      0.065     0.310    0.020     0.020
```

(cont.)

TABLE 5.8. *(cont.)*

```
    X10                0.083      0.065     1.282    0.083     0.081
    X11               -0.158      0.069    -2.305   -0.158    -0.154
    X12                0.000      0.000     0.000    0.000     0.000

 ENHANCE  BY
    X1                 0.000      0.000     0.000    0.000     0.000
    X2                -0.028      0.067    -0.410   -0.028    -0.027
    X3                -0.018      0.066    -0.263   -0.018    -0.017
    X4                -0.022      0.063    -0.353   -0.022    -0.021
    X5                -0.042      0.053    -0.789   -0.042    -0.043
    X6                 0.034      0.051     0.668    0.034     0.035
    X7                -0.039      0.054    -0.723   -0.039    -0.039
    X8                 0.000      0.000     0.000    0.000     0.000
    X9                 0.582      0.056    10.447    0.582     0.570
    X10                0.580      0.056    10.315    0.580     0.564
    X11                0.693      0.057    12.106    0.693     0.676
    X12                0.671      0.048    13.929    0.671     0.675

 SOCIAL   WITH
    COPING             0.400      0.100     3.983    0.400     0.400

 ENHANCE  WITH
    COPING             0.313      0.105     2.992    0.313     0.313
    SOCIAL             0.255      0.081     3.162    0.255     0.255

MODEL MODIFICATION INDICES
Minimum M.I. value for printing the modification index     10.000

                              M.I.     E.P.C.  Std E.P.C.  StdYX E.P.C.

WITH Statements
X10       WITH X9           24.856     0.213      0.213        0.203
X12       WITH X11          23.393     0.283      0.283        0.277
```

Although EFA may also furnish evidence of the presence of double loading items (see promax rotated loadings of item 4 in Table 5.6), EF[A] does not provide any direct indications of the potential existence of salien[t] correlated errors (as noted in Chapters 2 and 3, EFA identification restric[-] tions prevent the specification of correlated indicator errors). This i[s] another area where the results of E/CFA can be quite valuable. Because th[e] analysis is conducted within the CFA framework, modification indices an[d] other fit diagnostic information (e.g., standardized residuals) are availabl[e] to examine whether the observed correlations among indicators can b[e] adequately reproduced by the latent factors alone. Modification indices ar[e] provided only for the measurement error portion of the model (e.g., th[e] theta-delta matrix in LISREL) because all other portions of the solution ar[e] saturated (e.g., factor loadings and cross-loadings).

TABLE 5.7. *(cont.)*

```
std
  f1-f3 = 1.0,
  e1-e12 = td1-td12;
cov
  f1-f3 = ph21 ph31 ph32;
run;
```

Amos Basic 5

```
' Example of E/CFA in Amos 5.0
Sub Main ()
Dim sem As New AmosEngine
sem.TextOutput
sem.Standardized
sem.Smc

sem.BeginGroup "ecfa.txt"

sem.Structure "x1 =  COPING + (1) E1"                           ' X1 IS ANCHOR
sem.Structure "x2 =  COPING + SOCIAL + ENHANCE + (1) E2"
sem.Structure "x3 =  COPING + SOCIAL + ENHANCE + (1) E3"
sem.Structure "x4 =  COPING + SOCIAL + ENHANCE + (1) E4"
sem.Structure "x5 =  COPING + SOCIAL + ENHANCE + (1) E5"
sem.Structure "x6 =  COPING + SOCIAL + ENHANCE + (1) E6"
sem.Structure "x7 =  COPING + SOCIAL + ENHANCE + (1) E7"
sem.Structure "x8 =  SOCIAL + (1) E8"                           ' X8 IS ANCHOR
sem.Structure "x9 =  COPING + SOCIAL + ENHANCE + (1) E9"
sem.Structure "x10 = COPING + SOCIAL + ENHANCE + (1) E10"
sem.Structure "x11 = COPING + SOCIAL + ENHANCE + (1) E11"
sem.Structure "x12 = ENHANCE + (1) E12"                        ' X12 IS ANCHOR
sem.Structure "COPING (1)"
sem.Structure "SOCIAL (1)"
sem.Structure "ENHANCE (1)"
sem.Structure "COPING <--> SOCIAL"
sem.Structure "COPING <--> ENHANCE"
sem.Structure "SOCIAL <--> ENHANCE"

End Sub
```

Note. $N = 500$.

(e.g., see Table 4.1, Chapter 4), except for two major differences: (1) all factor loadings and cross-loadings are freely estimated, with the exception of the cross-loadings of the items selected as anchor indicators (X1, X6, X8); and (2) as in EFA, the metric of the latent factors is specified by fixing the factor variances to 1.0 (but the three-factor correlations are freely estimated). Note that items 1 (X1), 6 (X6), and 8 (X8) were selected as anchor indicators for Coping Motives, Social Motives, and Enhancement Motives, respectively. Anchor items are selected on the basis of EFA results (Table 5.6) and entail one item from each factor that has a high (or the highest) primary loading on the factor and low (or the lowest) cross-loadings on the remaining factors. For example, item 8 (X8) was chosen as the anchor indicator for Social Motives because it had the highest loading on this factor (.759) and the lowest cross-loadings on Coping Motives and Enhancement Motives (.011 and .001, respectively; see Table 5.6).

Selected Mplus output of the E/CFA is provided in Table 5.8. First, note that although this analysis was conducted in the CFA framework, the degrees of freedom and overall fit are the same as in the EFA solution presented in Table 5.6; $\chi^2(33) = 55.546$, $p = .008$, RMSEA = .037 (90% CI = .019 to .053, CFit = .898). Nonetheless, this analysis provides considerably more information than the EFA. The E/CFA provides z tests (labeled under the "Est./S.E." column in Mplus) to determine the statistical significance of primary and secondary loadings (except for the cross-loadings of the three anchor items).[8] For example, items 5 through 8 (X5–X8) have statistically significant ($ps < .001$) loadings on Social Motives (range of zs = 10.44 to 17.22), but none of these items have statistically significant cross-loadings on Coping Motives or Enhancement Motives (range of zs = 0.67 to 1.69). Conversely, whereas items 1 through 4 have statistically significant ($ps < .001$) loadings on Coping Motives, the E/CFA results indicated that item 4 also has a large (λ_{42} = .546) and statistically significant ($z = 8.21$, $p < .001$) loading on Social Motives. This suggests that a CFA model of congeneric indicator sets may not be viable; that is, item 4 should be specified to load on both Coping Motives and Social Motives, although its loading on Enhancement Motives could be fixed to zero (i.e., $z = 0.35$, Table 5.8). Other than item 4, the E/CFA results are generally supportive of fixing all other cross-loadings at zero in a subsequent CFA model. In addition, unlike the EFA, the E/CFA provides significance tests of the factor covariances; as seen in Table 5.8, the three drinking motives factors are significantly

TABLE 5.7. Syntax for Conducting an EFA within the CFA Framework for the Drinking Motives Questionnaire

LISREL 8.72

```
TITLE E/CFA OF DRINKING MOTIVES ITEMS
DA NI=13 NO=500 MA=CM
LA
X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 X11 X12 GP
RA FI = C:\EFA.SAV
SE
X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 X11 X12 /
MO NX=12 NK=3 PH=SY,FR LX=FU,FR TD=DI
LK
COPING SOCIAL ENHANCE
PA LX
1 0 0                          ! ITEM 1 IS AN ANCHOR ITEM FOR COPING
1 1 1
1 1 1
1 1 1
1 1 1
1 1 1
1 1 1
0 1 0                          ! ITEM 8 IS AN ANCHOR ITEM FOR SOCIAL
1 1 1
1 1 1
1 1 1
0 0 1                          ! ITEM 12 IS AN ANCHOR ITEM FOR ENHANCE
PA PH
0
1 0
1 1 0                            ! FACTOR COVARIANCES FREELY ESTIMATED
VA 1.0 PH(1,1) PH(2,2) PH(3,3)   ! FACTOR VARIANCES FIXED TO 1.0
OU ME=ML RS MI SC ND=4
```

Mplus 3.11

```
TITLE:  DRINKING MOTIVES ECFA
DATA:
  FILE IS "C:\efa.sav";
VARIABLE:
  NAMES ARE X1-X12 GP;
  USEV ARE X1-X12;
ANALYSIS: ESTIMATOR IS ML;
          ITERATIONS=1000;
MODEL:
  COPING BY  X1-X12*.5 X8@0 X12@0;  !X1 is ANCHOR ITEM
  SOCIAL BY  X1-X12*.5 X1@0 X12@0;  !X8 is ANCHOR ITEM
  ENHANCE BY X1-X12*.5 X1@0 X8@0;   !X12 is ANCHOR ITEM
  COPING-ENHANCE@1;                 !FACTOR VARIANCES FIXED TO 1.0
```


TABLE 5.7. *(cont.)*

EQS 5.7b

```
/TITLE
 E/CFA OF DRINKING MOTIVES ITEMS
/SPECIFICATIONS
 CASES=500; VAR=13; ME=ML; MA=RAW;
 DA = 'C:\EFA.SAV';
/LABELS
 v1=item1; v2= item2; v3= item3; v4= item4; v5= item5; v6= item6;
 v7= item7; v8= item8; v9= item9; v10=item10; v11=item11; v12=item12;
 v13=gp; f1 = coping; f2 = social; f3 = enhance;
/EQUATIONS
 V1 =  *F1+E1;              ! ITEM 1 IS ANCHOR ITEM FOR COPING
 V2 =  *F1+*F2+*F3+E2;
 V3 =  *F1+*F2+*F3+E3;
 V4 =  *F1+*F2+*F3+E4;
 V5 =  *F1+*F2+*F3+E5;
 V6 =  *F1+*F2+*F3+E6;
 V7 =  *F1+*F2+*F3+E7;
 V8 =  *F2+E8;              ! ITEM 8 IS ANCHOR ITEM FOR SOCIAL
 V9 =  *F1+*F2+*F3+E9;
 V10 = *F1+*F2+*F3+E10;
 V11 = *F1+*F2+*F3+E11;
 V12 = *F3+E12;             ! ITEM 12 IS ANCHOR ITEM FOR ENHANCE
/VARIANCES
 F1 TO F3 = 1.0;            ! FACTOR VARIANCES FIXED TO 1.0
 E1 TO E12 = *;
/COVARIANCES
 F1 TO F3 = *;              ! FACTOR COVARIANCES FREELY ESTIMATED
/PRINT
 fit=all;
/LMTEST
/END
```

SAS PROC CALIS 8.2

```
proc calis data=EFADATA cov method=ml pall pcoves;
var = X1-X12;
lineqs
 X1 =  lm11 f1 + e1,
 X2 =  lm21 f1 + lm22 f2 + lm23 f3 + e2,
 X3 =  lm31 f1 + lm32 f2 + lm33 f3 + e3,
 X4 =  lm41 f1 + lm42 f2 + lm43 f3 + e4,
 X5 =  lm51 f1 + lm52 f2 + lm53 f3 + e5,
 X6 =  lm61 f1 + lm62 f2 + lm63 f3 + e6,
 X7 =  lm71 f1 + lm72 f2 + lm73 f3 + e7,
 X8 =  lm82 f2 + e8,
 X9 =  lm91 f1 + lm92 f2 + lm93 f3 + e9,
 X10 = lm101 f1 + lm102 f2 + lm103 f3 + e10,
```

loadings of nonanchor items are freely estimated on each factor). Whereas this specification produces the same model fit as maximum likelihood EFA, the CFA estimation provides considerably more information, including the statistical significance of cross-loadings and the potential presence of salient error covariances. Thus, the researcher can develop a realistic measurement structure prior to moving into the more restrictive CFA framework (for applied examples of this approach, see Brown, White, & Barlow, 2005; Brown, White, Forsyth, & Barlow, 2004; Campbell-Sills, Liverant, & Brown, 2004). In addition, E/CFA can be used to bring other variables (i.e., predictors or distal outcomes of the factors) into an EFA-type solution, eliminating the need for factor scores (see Chapters 2 and 3 for an overview of the limitations of factor scores).

To illustrate this strategy, the data from the drinking motives questionnaire are again used. However, in this example, suppose that the psychometric development of this measure is in the early stages and the researcher has only a sense of the correct number of common factors and the hypothesized pattern of item–factor relationships based on theory-driven item development and preliminary exploratory research (which may have led to the elimination of some poorly behaved items). As the next exploratory step, the researcher conducts an EFA in a larger sample (i.e., the current sample with an $N = 500$) with the intent of verifying that a three-factor solution provides acceptable fit, and that the primary loadings of the items are generally in accord with prediction (e.g., items 1 through 4 have their highest loadings on the latent dimension of Coping Motives). At this stage of psychometric evaluation, use of CFA is premature. Although the researcher has a firm conceptual sense of this measure (i.e., number of factors, conjectured pattern of item–factor relationships, as supported by preliminary research), the initial EFA findings are limited in their ability to fully guide the CFA specification (e.g., reasonability of fixing all cross-loadings and error covariances to zero).

Table 5.6 presents Mplus syntax and selected results of an EFA using the $N = 500$ sample. As can be seen in this table, the three-factor EFA solution provided a good fit to the data, $\chi^2(33) = 55.546$, $p = .008$, RMSEA = .037 (90% CI = .019 to .053, CFit = .898). Although these results support the viability of a three-factor model, the researcher wishes to further explore the latent structure of this measure before specifying a CFA solution in an independent sample.

Table 5.7 provides the syntax from several programs (LISREL, Mplus,

TABLE 5.6. Mplus Syntax and Selected Results for an Exploratory Factor Analysis of a Drinking Motives Questionnaire

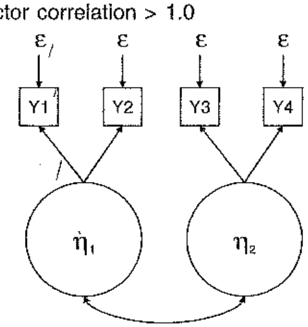
```
TITLE:  DRINKING MOTIVES EFA
DATA:
  FILE IS "C:\efa.sav";  ! the correlation matrix in Table 5.2 could be
VARIABLE:                ! used as input here
  NAMES ARE X1-X12 GP;
  USEV ARE X1-X12;
ANALYSIS:
  TYPE IS EFA 3 3;
  ESTIMATOR IS ML;
  ITERATIONS = 1000;
  CONVERGENCE = 0.00005;
OUTPUT: SAMPSTAT;

EIGENVALUES FOR SAMPLE CORRELATION MATRIX
              1          2          3          4          5
    1       3.876      1.906      1.150      0.837      0.722
              6          7          8          9         10
    1       0.669      0.576      0.557      0.487      0.471
             11         12
    1       0.426      0.323

EXPLORATORY ANALYSIS WITH  3 FACTOR(S) :
        CHI-SQUARE VALUE                 55.546
        DEGREES OF FREEDOM                   33
        PROBABILITY VALUE                0.0083
        RMSEA (ROOT MEAN SQUARE ERROR OF APPROXIMATION) :
        ESTIMATE (90 PERCENT C.I.) IS   0.037 ( 0.019  0.053)
        PROBABILITY RMSEA LE 0.05 IS    0.898

        PROMAX ROTATED LOADINGS
           1          2          3
X1       0.573     -0.037      0.031   *to be used as anchor item
X2       0.502      0.065      0.000
X3       0.475      0.071      0.009
X4       0.514      0.494      0.006
X5       0.095      0.590     -0.039
X6       0.107      0.671      0.041
X7       0.103      0.641     -0.034
X8      -0.011      0.759     -0.001   *to be used as anchor item
X9      -0.003      0.043      0.575
X10     -0.044      0.107      0.567
X11      0.108     -0.134      0.689
X12     -0.041      0.031      0.679   *to be used as anchor item
```

Model C: Positive Definite Input Matrix, Indefinite Model-Implied Matrix
(Specification problem: incorrect specification of indicator–factor relationships)

Input Matrix / Eigenvalues

	Y1	Y2	Y3	Y4	Eigenvalues
Y1	1.0000				2.501
Y2	0.4000	1.0000			0.851
Y3	0.7000	0.4000	1.0000		0.360
Y4	0.4500	0.6500	0.4000	1.0000	0.289

Consequence: Factor correlation > 1.0

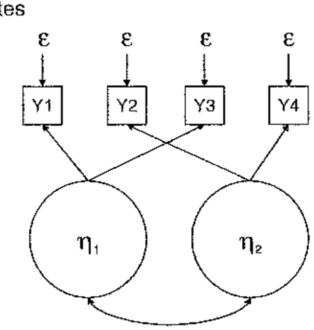


Model D: Positive Definite Input and Model-Implied Matrices
(Correctly specified model)

Input Matrix / Eigenvalues

	Y1	Y2	Y3	Y4	Eigenvalues
Y1	1.0000				2.501
Y2	0.4000	1.0000			0.851
Y3	0.7000	0.4000	1.0000		0.360
Y4	0.4500	0.6500	0.4000	1.0000	0.289

Consequence: Good-fitting model, χ^2 (1, N = 500) = 2.78, reasonable parameter estimates



might be addressed by respecifying the model with additional constraints (e.g., fixing the error variance to zero or a small positive value). Although overall fit will worsen somewhat, a proper solution may be obtained.[7] In fact, a default in the EQS program prevents negative variances by setting the lower bound of these estimates to zero. When the LISREL program encounters an indefinite or semidefinite matrix, it invokes a *ridge option*—a smoothing function to eliminate negative or zero eigenvalues—to make the input data suitable for the analysis. These remedial strategies are not recommended. As noted in this section, a negative indicator error variance may be due to a variety of problems, such as sampling fluctuation (small sample size), non-normality (e.g., outliers), multicollinearity, and model misspecification. In addition to causing the Heywood case, these issues signal other problems with the analysis and the sample data (e.g., low statistical power, poorly screened data, redundant indicators). Thus, it is better to diagnose and correct the true source of the problem rather than sidestep it with one of these quick fix remedies.

EFA IN THE CFA FRAMEWORK

A common sequence in scale development and construct validation is to conduct CFA as the next step after latent structure has been explored using EFA. However, the researcher frequently encounters a poor-fitting CFA solution because of the potential sources of misfit that are not present in EFA. For example, unlike the situation in EFA, indicator cross-loadings and residual covariances are usually fixed to zero in initial CFA models. The researcher is then faced with potentially extensive post hoc model testing subject to the criticisms of specification searches in a single data set (MacCallum, 1986).

Although underutilized in the applied literature, the procedure of "exploratory factor analysis within the CFA framework" (E/CFA; Jöreskog, 1969; Jöreskog & Sörbom, 1979; Muthén & Muthén, 1998) can be a useful precursor to CFA that allows the researcher to explore measurement structures more fully before moving into a confirmatory framework. The E/CFA approach represents an intermediate step between EFA and CFA that provides substantial information important in the development of realistic confirmatory solutions. In this strategy, the CFA applies the same number of identifying restrictions used in EFA (m^2) by fixing factor variances to unity, freely estimating the factor covariances, and by selecting an

In addition, the risk of nonconvergence and improper solutions is positively related to model complexity (i.e., models that have a large number of freely estimated parameters). A good example of an overparameterized model is the correlated methods CFA of multitrait–multimethod data (see Figure 6.1, Chapter 6). As discussed in Chapter 6, these models produce improper solutions most of the time. In other types of models, it may be possible to rectify the inadmissibility problem by removing some freely estimated parameters (e.g., if substantively justified, drop nonsignificant parameters or place other restrictions on the model, such as equality constraints; see Chapter 7). However, the estimation problems associated with model complexity are magnified by small sample size.

Some examples of nonpositive definite matrices and improper solutions are presented in Figure 5.2. In Model A, the input matrix is semidefinite because of a linear dependency in the sample data (i.e., Y1 is the sum of the other indicators). The fact that this matrix is not positive definite is upheld by the results of a PCA, which shows that one eigenvalue equals zero. Model B analyzes an indefinite input matrix; that is, one of its associated eigenvalues is negative. In this example, the correlation of Y1–Y2 cannot be –.20, given the relationships of Y1 and Y2 with other indicators in the sample data. This problem may have been produced by a data entry error or the use of pairwise deletion as the missing data strategy. Depending on the latent variable software program, Models A and B either will not be executed (i.e., the program will issue the warning that the matrix is not positive definite and will stop) or will produce a factor loading above 1.0 and a negative indicator error variance.

In Model C, the input matrix is positive definite: PCA indicates that all eigenvalues are positive. However, the two-factor measurement model is misspecified. The proposed model will not support the pattern of relationships in the sample data; for instance, Y1 and Y2 are specified as indicators of η_1, yet Y1 is much more strongly related to Y3 than Y2. In an effort to reproduce the sample relationships, given the model that is specified, the ML estimation process is "forced to push" one or more parameter estimates out of the range of admissibility. In this case, ML estimation produces an indefinite factor correlation matrix; that is, the correlation of η_1 and η_2 is estimated to be 1.368. When the model is properly specified (Model D), the results indicate a good-fitting model, for example, χ^2 (1, N = 500) = 2.78, p = .096, and reasonable parameter estimates; for example, factor correlation = .610, range of factor loadings = 0.778–0.867.

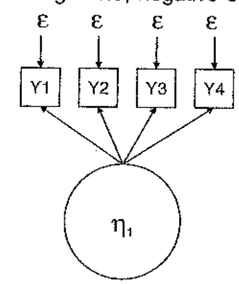
In practice, the problem of improper solutions is often circumvented by a "quick fix" method. F

Model A: Semidefinite Input Matrix
(Linear dependency: Y1 is the sum of Y2, Y3, and Y4)

Input Matrix

	Y1	Y2	Y3	Y4	Eigenvalue
Y1	1.0000				2.877
Y2	0.8035	1.0000			0.689
Y3	0.6303	0.4379	1.0000		0.435
Y4	0.8993	0.5572	0.3395	1.0000	0.000

Consequence: factor loading > 1.0, negative error variance

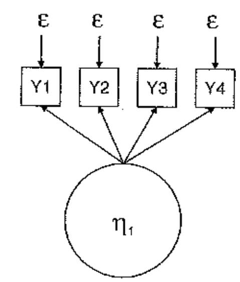


Model B: Indefinite Input Matrix
(Correlation of Y1–Y2 is out of possible range given other relationships)

Input Matrix

	Y1	Y2	Y3	Y4	Eigenvalues
Y1	1.0000				2.432
Y2	-0.2000	1.0000			1.209
Y3	0.7000	0.7000	1.0000		0.513
Y4	0.4500	0.6000	0.5000	1.0000	-0.155

Consequence: factor loading > 1.0, negative error variance



(cont

of these matrices must have determinants greater than zero ("principal submatrices" are all possible subsets of the original matrix created by removing variables from the original matrix).[5]

The condition of positive definiteness can be evaluated by submitting the variance–covariance matrix in question to principal components analysis (PCA; see Chapter 2). PCA will produce as many eigenvalues as the number of variables (p) in the input matrix. If all eigenvalues are greater than zero, the matrix is positive definite. If one or more eigenvalues are less than zero, the matrix is *indefinite*. The term *semidefinite* is used in reference to matrices that produce at least one eigenvalue that equals zero, but no negative eigenvalues (Wothke, 1993).

There are several potential causes of a nonpositive definite input variance–covariance matrix. As noted above, this problem may stem from high multicollinearities or linear dependencies in the sample data. This problem is usually addressed by eliminating collinear variables from the input matrix or combining them (cf. parceling, Chapter 9). Often, a nonpositive definite input matrix is due to a minor data entry problem, such as typographical errors in preparing the input matrix (e.g., a negative or zero sample variance, a correlation > 1.0) or errors in reading the data into the analysis (e.g., formatting errors in the syntax file). Large amounts of missing data, in tandem with use of an inferior approach to missing data management, can create a nonpositive definite input matrix. In Chapter 9, it is demonstrated that the range of possible values that a correlation (or covariance) may possess is dependent on all other relationships in the input matrix. For example, if $r_{x,z} = .80$ and $r_{y,z} = .80$, then $r_{x,y}$ must not be below .28 (see Eq. 9.1, Chapter 9) or this submatrix would not be positive definite. Input matrices created from complete data of a large sample are usually positive definite. However, pairwise deletion of missing data can cause definiteness problems (out-of-bound correlations) because the input matrix is computed on different subsets of the sample. Listwise deletion can produce a nonpositive definite matrix by decreasing sample size (see below) or other problems (e.g., creating constants in the sample data that have zero variance). Here, the best solution would be to use a state-of-the-art approach to missing data (i.e., direct ML, multiple imputation; see Chapter 9).

There are several other problems that may lead to improper solutions (i.e., nonpositive definite model matrices, Heywood cases). Perhaps the most common cause is a misspecified model: improper solutions frequently occur when the specified model is very different from models that the data would support. Structurally and empirically underidentified models will also lead to nonconverging or improper solutions (e.g., see Figure 3.7, Chapter 3; "Model Identification Revisited" section of this chapter). In these situations, it is often possible to revise the model using the fit diagnostic procedures described in this chapter. If the model is grossly misspecified, the researcher may need to move back into a purely exploratory analytic framework (i.e., EFA) to revamp the measurement model.[6] Bad starting values can be the root of improper solutions (see Chapter 3). However, with the possible exception of very complex models, starting values are rarely the cause of improper solutions in today's latent variable software programs, which have become very sophisticated in the automatic generation of such numbers.

Problems often arise with the use of small samples. For instance, an input matrix may not be positive definite owing to sampling error. Small samples often work in concert with other problems to create model estimation difficulties. For instance, small samples are more prone to the influence of outliers (i.e., cases with aberrant values on one or more variables). Outliers can cause collinearities and non-normality in the sample data and can lead to Heywood cases such as negative variance estimates (e.g., an indicator error less than zero). Moreover, some non-normal theory estimators (e.g., weighted least squares) perform very poorly with small or moderate-size samples because their associated weight matrices cannot be inverted (see Chapter 9). Anderson and Gerbing (1984) reported how sample size may interact with other aspects of a model to affect the risk for improper solutions. Specifically, these authors found that the risk of negative variance estimates is highest in small samples when there are only two or three indicators per latent variable and when the communalities of the indicators are low (see also Chen, Bollen, Paxton, Curran, & Kirby, 2001). As discussed in Chapter 10, each of these aspects also contributes to statistical power. In addition to reducing the risk of empirical underidentification, having multiple indicators per factor also decreases the likelihood of improper solutions. If an improper solution is caused by these issues, additional cases or data must be obtained; for example, collect a larger sample, or obtain a larger set of indicators with stronger relationships to the latent variables.

An estimator appropriate to the sample data must be used. For example, if the indicators are binary (e.g., Yes/No items), it would be inappropriate to conduct the CFA on a matrix of Pearson correlations or covariances using the ML estimator. This analysis would provide incorrect results. The proper procedures for conducting CFA with non-normal and categorical data are presented in Chapter 9.

the factor loadings for items 11 and 12 had to be increased to better approximate the observed relationship of these indicators, the factor loadings of items 9 and 10 were lowered in the iterations to better approximate the relationships of item 9 with items 11 and 12, and item 10 with items 11 and 12. However, these relationships were still overestimated by the parameter estimates (e.g., standardized residual for items 10 and 11 = –2.76, Table 5.5), and the reduction in the magnitude of the item 9 and item 10 factor loadings resulted in a model underestimate of the observed relationship of these indicators (i.e., standardized residual for items 9 and 10 = 4.48, Table 5.5). This again illustrates how correcting one misspecified parameter ($\delta_{12,11}$) may resolve several strains in the solution.

Because of the large sample sizes typically involved in CFA, the researcher will often encounter "borderline" modification indices (e.g., larger than 3.84, but not of particularly strong magnitude) that suggest that the fit of the model could be improved if correlated errors were added to the model. As with any type of parameter specification in CFA, correlated errors must be supported by a substantive rationale and should not be freely estimated simply to improve model fit. In addition, the magnitude of EPC values should also contribute to the decision about whether to free these parameters. As discussed in Chapter 4, the researcher should resist any temptation to use borderline modification indices to overfit the model. These trivial additional estimates usually have minimal impact on the key parameters of the CFA solution (e.g., factor loadings) and are apt to be highly unstable (i.e., reflect sampling error rather than an important relationship; cf. MacCallum, 1986).

It is also important to be consistent in the decision rules used to specify correlated errors; that is, if there is a plausible reason for correlating the errors of two indicators, then all pairs of indicators for which this reasoning applies should also be specified with correlated errors. For instance, if it is believed that method effects exist for questionnaire items that are reverse worded (e.g., Marsh, 1996), correlated errors should be freely estimated for all such indicators, not just a subset of them. Similarly, if the errors of indicators X1 and X2 and indicators X2 and X3 are correlated for the same reason, then the errors of X1 and X3 should also be estimated. Earlier in this chapter where the one- versus two-factor solution of Neuroticism and Extraversion was considered, it was shown that the patterning of standardized residuals and δ modification indices may suggest the existence of a distinct factor (see Table 5.1). However, in some situations, such as in the analysis of questionnaires with reverse-worded items (e.g., Brown, 2003; Marsh, 1996), this patterning may not reflect important latent dimensions but rather the impact of substantively irrelevant method effects. Theoretical considerations must strongly be brought to bear in this determination, as this decision may have far-reaching implications for the future measurement and conceptualization of the construct.

Improper Solutions and Nonpositive Definite Matrices

A measurement model should not be deemed acceptable if the solution contains one or more parameter estimates that have out-of-range values. As noted earlier, such estimates are usually referred to as *Heywood cases* or *offending estimates*. The most common type of Heywood case in a CFA model is a negative error variance. Moreover, a completely standardized factor loading with a value greater than 1.0 is problematic if the CFA consists of congeneric indicator sets (i.e., each indicator loads on one factor only). In CFA models with indicators that load on more than one factor (and the factors are specified to be intercorrelated), the factor loadings of such indicators are regression coefficients, *not* correlations between the indicators and the factors. A completely standardized factor loading above 1.0 may be admissible in such models, although this result might be indicative of multicollinearity in the sample data (see below). This section discusses the various sources of improper solutions and their remedies.

A necessary condition for obtaining a proper CFA solution is that both the input variance–covariance matrix and the model-implied variance–covariance matrix are *positive definite*. Appendix 3.3 (Chapter 3) introduced the concept of positive definiteness in context of the calculation of determinants and F_{ML}. As noted in Appendix 3.3, a determinant is a single number (*scalar*) that conveys the amount of nonredundant variance in a matrix (i.e., the extent to which variables in the matrix are free to vary). When a determinant equals 0, the matrix is said to be *singular*, meaning that one or more rows or columns in the matrix are linearly dependent on other rows and columns. An example of a singular matrix is one comprised of three test scores: Subscale A, Subscale B, and Total Score (sum of Subscale A and Subscale B). The resulting matrix is singular because the third variable (Total Score) is redundant; that is, it is linearly dependent on the other two variables and thus has no freedom to vary. A singular matrix is problematic because it has no inverse. Consequently, multivariate statistics that require the inverse of a matrix cannot be computed (e.g., F_{ML}; cf. Appendix 3.3). Singularity is one reason why a matrix will not be positive definite. For the condition of positive definiteness to hold, the input and model-implied matrices and every principal submatrix

TABLE 5.5. *(cont.)*

Modification Indices for THETA-DELTA

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
X1	- -											
X2	1.96	- -										
X3	2.04	0.04	- -									
X4	0.77	3.02	1.09	- -								
X5	0.24	2.50	0.17	0.34	- -							
X6	0.19	0.51	0.68	0.03	0.50	- -						
X7	0.02	0.47	0.31	1.05	0.26	2.40	- -					
X8	0.36	0.01	3.07	0.07	0.00	0.03	1.17	- -				
X9	0.15	1.29	0.02	0.54	0.85	0.98	0.02	1.80	- -			
X10	0.10	0.47	3.82	1.13	0.29	0.52	1.59	1.26	20.06	- -		
X11	2.02	0.34	3.49	1.31	1.02	0.34	0.18	1.79	3.65	7.60	- -	
X12	1.91	0.97	0.08	0.32	1.22	1.52	0.58	0.59	7.21	4.50	**26.02**	- -

Completely Standardized Expected Change for THETA-DELTA

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
X1	- -											
X2	0.06	- -										
X3	-0.06	-0.01	- -									
X4	0.03	-0.07	0.04	- -								
X5	-0.02	0.05	0.01	-0.02	- -							
X6	-0.01	0.02	0.03	0.00	0.02	- -						
X7	0.00	-0.02	0.02	0.03	-0.02	-0.05	- -					
X8	-0.02	0.00	-0.05	-0.01	0.00	0.01	0.03	- -				
X9	0.01	0.04	-0.01	-0.02	0.03	-0.03	0.00	0.04	- -			
X10	0.01	-0.02	-0.07	0.03	0.02	0.02	-0.04	0.03	0.18	- -		
X11	0.05	0.02	0.06	-0.03	-0.03	0.02	-0.01	-0.04	-0.08	-0.12	- -	
X12	-0.05	-0.03	-0.01	0.01	-0.03	0.03	0.02	-0.02	-0.12	-0.09	**0.25**	- -

Completely Standardized Solution

LAMBDA-X

	Coping	Social	Enhance
X1	0.5137	- -	- -
X2	0.5149	- -	- -
X3	0.5160	- -	- -
X4	0.5380	0.4390	- -
X5	- -	0.6319	- -
X6	- -	0.7465	- -
X7	- -	0.6906	- -
X8	- -	0.7309	- -
X9	- -	- -	0.6026
X10	- -	- -	0.5947
X11	- -	- -	0.6649
X12	- -	- -	0.6629

PHI

	Coping	Social	Enhance
Coping	1.0000		
Social	0.6096	1.0000	
Enhance	0.3500	0.2645	1.0000

THETA-DELTA

X1	X2	X3	X4	X5	X6
0.74	0.73	0.73	0.23	0.60	0.44

X7	X8	X9	X10	X11	X12
0.52	0.47	0.64	0.65	0.56	0.56

errors may be needed to account for method covariance, such as in the analysis of indicators collected from different assessment modalities (e.g., self-report, behavioral observation, interview rating; cf. the multitrait–multimethod approach, Chapter 6).

Unnecessary correlated errors can be readily detected by results indicating their statistical or clinical nonsignificance (e.g., z values below 1.96, or very small parameter estimates that reflect trivial shared variance of the errors). The next step is simply to refit the model by fixing these covariances to zero and verify that this respecification does not result in a significant decrease in model fit. The χ^2_{diff} test can be used in this situation (but see below). The more common difficulty is the failure to include salient correlated errors in the solution. As in the prior examples, the omission of these parameters is typically manifested by large standardized residuals, modification indices, and EPC values.

Table 5.5 presents selected results of the drinking motives CFA solution where the correlated error between items 11 and 12 has not been specified (all other aspects of the solution were properly specified). Although the overall fit of the model is good, for example, $\chi^2(50) = 69.30$, $p = .037$, CFI = .99, RMSEA = .029, SRMR = .031, standardized residuals (5.10) and modification indices ($\delta_{12,11} = 26.02$) indicate that the relationship between these items has not been adequately reproduced by the model's parameter estimates. The need for a correlated error for these items is further evidenced by a rather large completely standardized expected change value ($\delta_{12,11} = .25$). Because this correlated error can be defended substantively (i.e., item 11 and item 12 are the only reverse-worded items in this questionnaire), this parameter is freed in the respecified solution. This modification significantly improves model fit, $\chi^2_{\text{diff}}(1) = 24.43$, $p < .001$, and the completely standardized estimate of this correlated error is .214 (see Figure 5.1).

As can be seen in Table 5.5, consequences of this misspecification include higher factor loadings for items 11 and 12, and lower loadings for items 9 and 10, on the Enhancement Motives factor. The factor loadings of items 11 and 12 were inflated in the iterative process because a considerable portion of the observed correlation between these indicators could not be reproduced by the correlated error; for example, $\lambda_{x11,3} = .665$ versus .542 in the Table 5.5 and Figure 5.1 solutions, respectively. However, in attempt to avoid marked overestimation of the relationships between items 9 and 10 with items 11 and 12, the magnitude of the item 9 and item 10 factor loadings is attenuated; for example, $\lambda_{x10,3} = .595$ versus .664 in the Table 5.5 and Figure 5.1 solutions, respectively. In other words, because

TABLE 5.5. Selected Results for the Drinking Motives Three-Factor Model without the Correlated Error of Item 11 and Item 12

Standardized Residuals

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
X1	- -											
X2	1.40	- -										
X3	-1.43	-0.19	- -									
X4	-0.34	-1.07	1.61	- -								
X5	-0.76	1.58	0.56	-0.22	- -							
X6	-0.65	1.11	1.06	0.56	0.71	- -						
X7	-0.50	-0.09	0.74	0.95	-0.51	-1.55	- -					
X8	-1.42	0.02	-1.45	-1.27	0.07	0.18	1.08	- -				
X9	0.67	0.97	-0.02	0.08	0.56	0.34	-0.06	0.97	- -			
X10	0.70	-0.21	-1.20	1.36	0.88	2.00	-0.10	1.54	4.48	- -		
X11	1.09	0.30	1.02	-1.51	-1.84	-0.22	-1.59	-2.01	-1.91	-2.76	- -	
X12	-0.86	-0.85	-0.50	-0.14	-0.93	1.39	0.20	-0.48	-2.69	-2.12	5.10	- -

Modification Indices for LAMBDA-X

	Coping	Social	Enhance
X1	- -	1.94	0.31
X2	- -	0.72	0.00
X3	- -	0.30	0.04
X4	- -	- -	0.11
X5	0.08	- -	0.46
X6	1.41	- -	2.70
X7	0.24	- -	0.47
X8	3.66	- -	0.09
X9	0.24	0.31	- -
X10	0.56	3.16	- -
X11	0.12	4.92	- -
X12	0.60	0.02	- -

Completely Standardized Expected Change for LAMBDA-X

	Coping	Social	Enhance
X1	- -	-0.12	0.03
X2	- -	0.07	0.00
X3	- -	0.05	-0.01
X4	- -	- -	-0.01
X5	0.02	- -	-0.03
X6	0.09	- -	0.07
X7	0.04	- -	-0.03
X8	-0.14	- -	-0.01
X9	0.03	0.03	- -
X10	0.04	0.08	- -
X11	-0.02	-0.10	- -
X12	[illegible]	[illegible]	- -

where a is the number of freely estimated parameters in the model (in Figure 5.1, a = 29). Using this equation, AIC would equal 102.87 (44.87 + 58). Mplus output presents the AIC as

$$\text{AIC} = -2(\text{loglikelihood}) + 2a \tag{5.3}$$

For a demonstration of the equivalence of these three equations, see Kaplan (2000).

Although Mplus and EQS do not provide the ECVI (the ECVI is calculated in LISREL and Amos), its approximation is straightforward:

$$\text{ECVI} = (\chi^2/n) + 2(a/n) \tag{5.4}$$

where $n = N - 1$ and a is the number of freely estimated parameters in the model. Thus, the ECVI for the Figure 5.1 solution is .21; that is, (44.8654/499) + 2(29/499) = .09 + .12 = .21. For the current misspecified model involving the item 12 loading, AIC and ECVI are 184.34 and .37, respectively (AIC calculated using Eq. 5.2).

Whereas the AIC and ECVI can be used in tandem with χ^2_{diff} for the comparison of nested solutions, these indices are more often considered in the evaluation of competing, non-nested models. Generally, models with the lowest AIC and ECVI values are judged to fit the data better in relation to alternative solutions (regardless of the method by which AIC is calculated). From this standpoint, the Figure 5.1 solution would be favored over the current model because it is associated with lower AIC and ECVI values (although in the present case, this could be determined simply by comparing model χ^2s because the two solutions do not differ in *df*). It must be emphasized that, unlike χ^2_{diff}, the AIC and ECVI do not provide a *statistical* comparison of competing models. Rather, these indices foster the comparison of the overall fit of models, adjusting for the complexity of each.[4]

Another possible problematic outcome of a CFA solution is that an indicator does not load on any latent factor. This problem is readily diagnosed by results showing that the indicator has a nonsignificant or nonsalient loading on the conjectured factor, as well as modification indices (and EPC values) suggesting that the fit of the model could not be improved by allowing the indicator to load on a different factor. This conclusion should be further supported by inspection of standardized residuals (2.00) and sample correlations that point to the fact that the indicator is weakly related to other indicators in the model. Unlike the prior exam-

ples, this scenario does not substantially degrade the fit of the moc (assuming that the model is well specified otherwise). Thus, although t proper remedial action is to eliminate the problematic indicator, the ove all fit of the model will usually not improve. In addition, the revis model is not nested with the initial solution because the input varianc covariance matrix has changed.

Although most of this section has described various ways a moc might be respecified retaining the original set of indicators, it is often t case that a much better fitting solution can be obtained by simply dro ping bad indicators from the model. For example, an indicator may associated with several large modification indices and standardized resid als, reflecting that the indicator is rather nonspecific in that it evidenc similar relationships to all latent variables in the solution. Simply droppi this indicator from the model will eliminate multiple strains in the sol tion.

Correlated Errors

A CFA solution can also be misspecified with respect to the relationshi among the indicator error variances. When no correlated errors (er covariances) are specified, the researcher is asserting that all of t covariation among indicators that load on a given factor is due to th latent dimension and that all measurement error is random. Correlat errors between indicators are specified on the basis of the notion that so of the covariance in the indicators not explained by the latent variable due to another exogenous common cause; that is, some of the shared va ance is due to the latent factor, some of the shared variance is due to outside cause. It is also possible that correlated errors exist for indicat that load on separate latent factors. In this case, most of the shared va ance may be due to an outside cause (some of the observed covariati may also be reproduced by the product of the indicators' factor loadir and the factor correlation). Occasionally, variables measured by sin indicators are included in a CFA model (see Chapters 4 and 7). Becau such models would be underidentified, it is not possible to correlate t measurement errors of single indicators with the errors of other indicat in the solution.

As discussed earlier in this chapter, in the case of the analysis of m tiple questionnaire items, correlated errors may arise from items that very similarly worded, reverse-worded, or differentially prone to soc desirability, and so forth. In CFA construct validation studies, correlat

TABLE 5.4. *(cont.)*

Completely Standardized Solution

LAMBDA-X

	Coping	Social	Enhance
X1	0.5163	- -	- -
X2	0.5178	- -	- -
X3	0.5158	- -	- -
X4	0.5301	0.4468	- -
X5	- -	0.6297	- -
X6	- -	0.7471	- -
X7	- -	0.6886	- -
X8	- -	0.7293	- -
X9	- -	- -	0.6683
X10	- -	- -	0.6599
X11	- -	- -	0.3816
X12	- -	0.2048	- -

PHI

	Coping	Social	Enhance
Coping	1.0000		
Social	0.6084	1.0000	
Enhance	0.3524	0.3222	1.0000

THETA-DELTA

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
X1	0.73											
X2	- -	0.73										
X3	- -	- -	0.73									
X4	- -	- -	- -	0.23								
X5	- -	- -	- -	- -	0.60							
X6	- -	- -	- -	- -	- -	0.44						
X7	- -	- -	- -	- -	- -	- -	0.53					
X8	- -	- -	- -	- -	- -	- -	- -	0.47				
X9	- -	- -	- -	- -	- -	- -	- -	- -	0.55			
X10	- -	- -	- -	- -	- -	- -	- -	- -	- -	0.56		
X11	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	0.85	
X12	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	0.40	0.96

TABLE 5.4. Selected Results for the Drinking Motives Three-Factor Model When Item 12 Has Been Specified to Load on the Wrong Factor

Standardized Residuals

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	- -											
2	1.32											
3	-1.49	-0.24	- -									
4	-0.36	-1.13	1.89	- -								
5	-0.76	1.57	0.60	-0.21	- -							
6	-0.68	1.08	1.07	0.37	0.77	- -						
7	-0.50	-0.09	0.77	0.94	-0.37	-1.47	- -					
8	-1.42	0.01	-1.42	-1.30	0.20	0.23	1.22	- -				
9	0.35	0.68	-0.38	-1.61	-0.40	-1.01	-1.26	-0.20	- -			
10	0.39	-0.60	-1.64	0.21	-0.03	0.97	-1.28	0.49	1.54	- -		
11	2.35	1.62	2.30	1.15	-0.77	0.96	-0.40	-0.66	6.07	5.18	8.78	
12	0.55	0.57	0.88	1.48	-1.54	0.85	-0.41	-1.26	7.32	7.40	7.37	0.41

Modification Indices for LAMBDA-X

	Coping	Social	Enhance
1	- -	1.98	1.03
2	- -	0.67	0.29
3	- -	0.34	0.28
4	- -	- -	0.64
5	0.13	- -	0.09
6	0.78	- -	0.09
7	0.18	- -	2.60
8	3.61	- -	0.00
9	0.98	1.21	- -
10	0.34	0.67	- -
11	4.59	0.17	- -
12	1.11	- -	**77.01**

Completely Standardized Expected Change for LAMBDA-X

	Coping	Social	Enhance
1	- -	-0.12	0.06
2	- -	0.07	0.03
3	- -	0.05	-0.03
4	- -	- -	-0.04
5	0.03	- -	-0.02
6	0.06	- -	0.01
7	0.03	- -	-0.08
8	-0.14	- -	0.00
9	-0.07	-0.07	- -
10	-0.04	0.05	- -
11	0.12	0.02	- -
12	0.08	- -	**0.53**

Modification Indices for THETA-DELTA

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	- -											
2	1.73	- -										
3	2.23	0.06	- -									
4	0.81	3.16	1.56	- -								
5	0.26	2.44	0.17	0.39	- -							
6	0.15	0.60	0.83	0.03	0.60	- -						
7	0.01	0.43	0.37	1.15	0.14	2.15	- -					
8	0.37	0.02	2.95	0.12	0.04	0.05	1.48	- -				
9	0.03	1.05	0.02	1.78	0.03	1.31	0.05	0.26	- -			
10	0.01	1.19	3.84	0.69	0.07	0.72	2.30	0.07	2.37	- -		
11	2.77	0.76	3.97	0.20	0.30	0.66	0.04	0.68	0.00	1.34	- -	
12	0.59	0.38	0.53	0.36	1.85	0.12	0.20	1.15	20.96	22.72	- -	- -

ters may take on out-of-range values to minimize F_{ML}. In this example, a solution with Heywood cases can be produced by fitting the Figure 5.1 model, by specifying X7 to load on Coping Motives instead of Social Motives. Also see in Table 5.3 that the standardized residuals are not clearly diagnostic of the source of misspecification in this instance; for instance, these values suggest that the relationships between item 4 and the Social Motives indicators have been reproduced adequately. This is because the ML iterations produced parameter estimates that were able to reproduce most relationships reasonably well. Focal areas of strains are nonetheless evident; e.g., some relationships are overestimated (e.g., standardized residual for X3, X8 = –2.56), and one relationship is underestimated (standardized residual for X1, X2 = 3.22), each relating to strains in the model for reproducing the relationship of the Coping Motives and Social Motives indicators. In addition to emphasizing the need to examine multiple aspects of fit (i.e., overall fit, standardized residuals, modification indices, parameter estimate values), this illustrates that respecifying one parameter may successfully eliminate what seem to be multiple strains, because all relationships in the model are interdependent.

Moreover, an indicator–factor relationship may be misspecified when an indicator loads on the wrong factor. Table 5.4 presents selected results of an analysis of the drinking motives model in which item 12 was specified to load on the Social Motives factor (its relationship to Enhancement Motives was fixed to 0, but its correlated error with item 11 was freely estimated). As in the prior example, overall fit statistics suggest acceptable model fit, $\chi^2(49) = 126.337$, $p < .001$, SRMR = .063, RMSEA = .054 (90% CI = 0.042 to 0.066; CFit $p = .302$), TLI = .954, CFI = .971, based on the guidelines recommended by Hu and Bentler (1999). However, as shown in Table 5.4, modification indices and standardized residuals indicate localized points of strain in the solution. As before, simply freeing the fixed parameter ($\lambda_{12,3}$) with the largest modification index (77.01) would result in the proper measurement model. This misspecification is evident in other ways, such as (1) large standardized residuals indicating that the observed relationships among the Enhancement Motives indicators (e.g., item 12 with items 9–11) are being underestimated by the model parameter estimates; and (2) a factor loading of item 12 with Social Motives ($\lambda_{12,2}$ = .20) that, while statistically significant (z = 4.46, not shown in Table 5.4), is well below conventional guidelines for a "salient" indicator–factor relationship. Note that another impact of this misspecification is an elevated model estimate of the correlated error of items 11 and 12 ($\delta_{12,11}$ = .40 [illegible]

reflects the attempts of the iterations to reproduce the observed correlation between items 11 and 12 (r = .507), although the model still underestimates this relationship (standardized residual = 7.37). In the correct solution, the sample correlation is reproduced by the sum of the product of the item 11 and item 12 factor loadings [$\lambda_{11,3}\lambda_{12,3}$ = .542 (.541) = .293] and the correlated error of these indicators; that is, .293 + .214 = .507. Because the misspecified model cannot use the product of the item 11 and item 12 factor loadings (because item 12 loads on a different factor than item 11), the solution must rely more on the correlated error ($\delta_{12,11}$) to reproduce this observed relationship. A very small portion of this relationship is also estimated in the solution by $\lambda_{12,2}\phi_{32}\lambda_{11,3}$.

Unlike some previous examples, the Figure 5.1 model and the current model are not nested. Both models entail 29 freely estimated parameters (df = 49), and thus one does not contain a subset of the freed parameters of the other. Therefore, the χ^2_{diff} test cannot be used to statistically compare these two solutions. In this scenario, a strategy that can be used to compare solutions is to qualitatively evaluate each with regard to the three major aspects of model acceptability: overall goodness of fit, focal areas of ill fit, and interpretability/strength of parameter estimates. For instance, if one model satisfies each of these criteria and the other does not, the former model would be favored.

In addition, methodologists have developed procedures for using χ^2 in the comparison of non-nested models. Two popular methods are the Akaike Information Criterion (AIC; Akaike, 1987) and the Expected Cross-Validation Index (ECVI; Browne & Cudeck, 1989). These indices are closely related in that they both take into account model fit (as reflected by χ^2) and model complexity/parsimony (as reflected by model df or the number of freely estimated parameters; cf. the RMSEA). The ECVI also incorporates sample size—specifically, a greater penalty function for fitting a nonparsimonious model in a smaller sample.

Latent variable software programs differ in their computation of the AIC. For instance, in EQS, the AIC is computed as follows:

$$\text{AIC} = \chi^2 - 2df \tag{5.1}$$

Thus, for the Figure 5.1 model, AIC would be –53.13 (i.e., 44.8654 – 98). In LISREL and Amos, AIC is computed as

$$\text{AIC} = \chi^2 + 2a \tag{5.2}$$

TABLE 5.3. *(cont.)*

Completely Standardized Solution

LAMBDA-X

	Coping	Social	Enhance
X1	0.4311	- -	- -
X2	0.4349	- -	- -
X3	0.4502	- -	- -
X4	0.9548	- -	- -
X5	- -	0.6333	- -
X6	- -	0.7478	- -
X7	- -	0.6895	- -
X8	- -	0.7291	- -
X9	- -	- -	0.6663
X10	- -	- -	0.6688
X11	- -	- -	0.5370
X12	- -	- -	0.5414

PHI

	Coping	Social	Enhance
Coping	1.0000		
Social	0.7978	1.0000	
Enhance	0.3348	0.2917	1.0000

substantively), the correct model can be readily obtained. As seen in Table 5.3, the fixed parameter with by far the highest modification index is λ_{42} (18.04), corresponding to a double-loading of item 4 with the Social Motives factor (cf. Figure 5.1).

Second, these results demonstrate that the acceptability of the model should not be based solely on indices of overall model fit, although this practice is somewhat common in the literature. Note that all descriptive fit indices were consistent with good model fit. However, these global indices masked the fact that at least one relationship was not well represented by the model (i.e., the double-loading of item 4).

Third, the modification index (18.04) and completely standardized EPC values (1.00) do not correspond exactly to the actual change in model χ^2 and parameter estimate, respectively, when the relationship between item 4 and the Social Motives factor is freely estimated. As seen in Figure 5.1, the completely standardized parameter estimate of this double-loading is .438. This revision produces a significant improvement in model fit, $\chi^2_{\text{diff}}(1) = 16.67$, $p < .01$ (i.e., $61.535 - 44.865 = 16.67$), but the χ^2 difference is slightly smaller than the modification index associated with this parameter (18.04). Again, this underscores the fact that modification indices and EPC values are approximations of model change if a fixed or constrained parameter is freed.

Fourth, aspects of these results illustrate that parameter estimates should not be interpreted when the model is poor fitting.[3] For instance, it can be seen in Table 5.3 that some of the consequences of failing to specify the item 4 cross-loading include an inflated estimate of the factor correlation between Coping Motives and Social Motives (.798), an inflated estimate of the loading of item 4 on Coping Motives (.955), and underestimates of the loadings of items 1–3 on Coping Motives. Because item 4 was not specified to load on Social Motives, its moderate relationships with the Social Motives indicators (items 5–8, *r*s = .48 to .58; see input correlation matrix in Table 5.2) had to be reproduced primarily through its factor loading with Coping Motives (λ_{41} = .955) and the factor correlation of Coping and Social Motives (ϕ_{21} = .798); e.g., the model-predicted relationship of X4 and X5 (completely standardized) = $\lambda_{41}\phi_{21}\lambda_{52}$ = .955(.798)(.633) = .482 (observed *r* of X4 and X5 was .481; see Table 5.2). Thus, as the result of the iterative process of minimizing F_{ML}, these parameters (λ_{41}, ϕ_{21}) were inflated (and λ_{11}, λ_{21}, and λ_{31} were underestimated) in order to acquire a model that best reproduced the observed relationships ($S - \Sigma$). This also exemplifies why Heywood cases (offending estimates) may arise from a misspecified solution—through the iterative process, the parame-

TABLE 5.3. Selected Results for the Drinking Motives Three-Factor Model When a Double-Loading Item Has Not Been Specified

Standardized Residuals

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
X1	- -											
X2	3.22	- -										
X3	1.01	1.89	- -									
X4	-0.17	-2.65	-0.26	- -								
X5	-1.36	0.96	-0.29	-0.08	- -							
X6	-1.44	0.29	-0.05	0.80	0.60	- -						
X7	-1.16	-0.79	-0.21	1.24	-0.53	-1.57	- -					
X8	-2.15	-0.75	-2.56	-0.82	0.07	0.21	1.20	- -				
X9	0.94	1.22	0.19	-1.14	-0.03	-0.45	-0.77	0.29	- -			
X10	0.94	0.04	-1.00	0.39	0.27	1.34	-0.88	0.86	0.36	- -		
X11	1.99	1.27	1.87	-0.35	-1.43	0.19	-1.12	-1.46	0.86	-0.64	- -	
X12	0.21	0.22	0.47	0.81	-0.60	1.64	0.51	-0.09	-0.34	-0.22	- -	- -

Modification Indices for LAMBDA-X

	Coping	Social	Enhance
X1	- -	6.63	1.89
X2	- -	0.03	0.89
X3	- -	1.92	0.02
X4	- -	**18.04**	2.16
X5	0.03	- -	0.19
X6	0.44	- -	1.37
X7	0.81	- -	1.06
X8	1.94	- -	0.01
X9	0.76	0.35	- -
X10	0.18	0.53	- -
X11	0.16	1.53	- -
X12	0.77	1.16	- -

Completely Standardized Expected Change for LAMBDA-X

	Coping	Social	Enhance
X1	- -	-0.27	0.07
X2	- -	-0.02	0.05
X3	- -	-0.15	0.01
X4	- -	1.00	-0.10
X5	-0.01	- -	-0.02
X6	0.06	- -	0.05
X7	0.08	- -	-0.05
X8	-0.13	- -	0.01
X9	-0.04	-0.03	- -
X10	0.02	0.04	- -
X11	-0.02	-0.06	- -
X12	0.04	0.05	- -

Modification Indices for THETA-DELTA

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
X1	- -											
X2	10.37	- -										
X3	1.02	3.58	- -									
X4	0.03	7.04	0.07	- -								
X5	0.45	1.45	0.03	0.10	- -							
X6	0.47	0.11	0.23	0.08	0.37	- -						
X7	0.28	1.12	0.01	2.22	0.28	2.45	- -					
X8	1.53	0.36	5.21	0.14	0.00	0.04	1.43	- -				
X9	0.50	2.18	0.27	2.59	0.45	1.28	0.01	0.93	- -			
X10	0.04	0.70	3.39	0.49	0.10	0.83	1.71	0.68	0.13	- -		
X11	4.98	2.06	5.92	1.37	0.72	0.12	0.41	1.58	0.67	0.24	- -	
X12	1.93	0.87	0.32	1.14	0.74	1.53	0.68	0.23	0.26	0.00	- -	- -

(cont.)

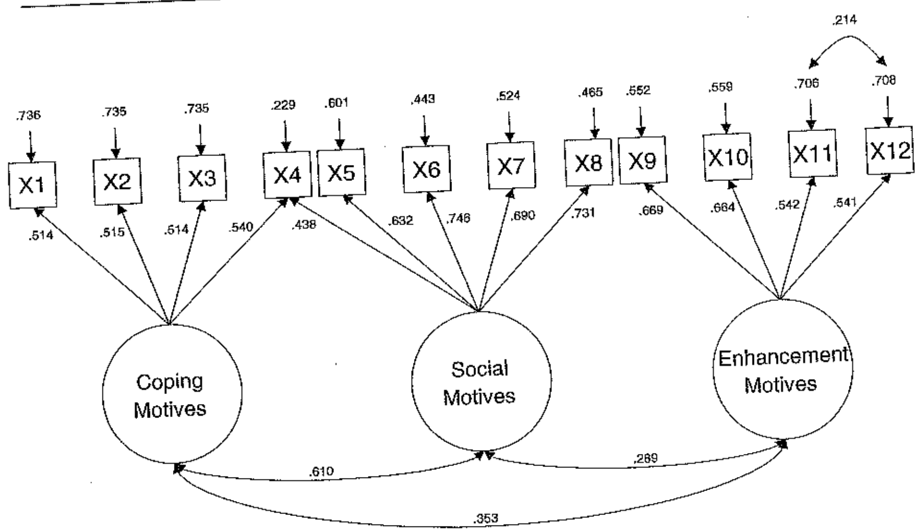


FIGURE 5.1. Completely standardized parameter estimates from the three-factor CFA model of a 12-item Alcohol Drinking Motives Questionnaire (N = 500). Overall fit of the model: $\chi^2(49)$ = 44.865, p = .641, SRMR = .025, RMSEA = 0.00 (90% CI = 0.00 to 0.025; CFit p = 1.00) TLI = 1.002, CFI = 1.00. All freely estimated unstandardized parameter estimates are statistically significant (p < .001).

dimensions of drinking motives (including item 4, which is expected to load only on the Social Motives factor). When this model is fit to the data (the correlated error of items 11 and 12 is included), the following fit indices result: $\chi^2(50)$ = 61.535, p = .127, SRMR = .032, RMSEA = .025 (90% CI = 0.00 to .040; CFit p = .998), TLI = .994, CFI = .996. Note that this solution has one more df than the "true" model because the relationship between item 4 and Social Motives has been fixed to zero; hence, this solution is nested under the "true" model.[2]

Selected results of this solution are presented in Table 5.3: standardized residuals, modification indices, completely standardized expected parameter change values (EPCs), and completely standardized estimates of factor loadings and factor correlations. These data exemplify several points made earlier in this chapter and in Chapter 4. For instance, researchers (e.g., MacCallum, 1986) have found that specification searches based on modification indices are more likely to be successful when the model contains only minor misspecifications. Thus, following the recommendation (cf. Jöreskog, 1993) of freeing a fixed or constrained parameter with the ... modification index (provided that this parameter can be interpreted

TABLE 5.2. LISREL Syntax for a Three-Factor Model of a Alcohol Drinking Motives Questionnaire

```
TITLE THREE FACTOR MODEL FOR DRINKING MOTIVES
DA NI=12 NO=500 MA=CM
LA
X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 X11 X12
KM
1.000
0.300 1.000
0.229 0.261 1.000
0.411 0.406 0.429 1.000
0.172 0.252 0.218 0.481 1.000
0.214 0.268 0.267 0.579 0.484 1.000
0.200 0.214 0.241 0.543 0.426 0.492 1.000
0.185 0.230 0.185 0.545 0.463 0.548 0.522 1.000
0.134 0.146 0.108 0.186 0.122 0.131 0.108 0.151 1.000
0.134 0.099 0.061 0.223 0.133 0.188 0.105 0.170 0.448 1.000
0.160 0.131 0.158 0.161 0.044 0.124 0.066 0.061 0.370 0.350 1.000
0.087 0.088 0.101 0.198 0.077 0.177 0.128 0.112 0.356 0.359 0.507 1.000
SD
2.06 1.52 1.92 1.41 1.73 1.77 2.49 2.27 2.68 1.75 2.57 2.66
MO NX=12 NK=3 PH=SY,FR LX=FU,FR TD=SY,FR
LK
Coping Social Enhance
PA LX
0 0 0
1 0 0
1 0 0
1 1 0     !double loading item
0 0 0
0 1 0
0 1 0
0 1 0
0 0 0
0 0 1
0 0 1
0 0 1
VA 1.0 LX(1,1) LX(5,2) LX(9,3)
PA TD
1
0 1
0 0 1
0 0 0 1
0 0 0 0 1
0 0 0 0 0 1
0 0 0 0 0 0 1
0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 0 0 1 1     !correlated residual
PA PH
1
1 1
1 1 1
OU ME=ML RS MI SC AD=OFF IT=100 ND=4
```

tific evidence for the distinctiveness of Neuroticism and Extraversion. Thus, the researcher should provide justification for both the hypothesized and alternative CFA models.

If too many factors have been specified in the CFA model, this is likely to be detected by correlations between factors that approximate ±1.0, and so the latent dimensions have poor *discriminant validity.* In applied research, a factor correlation that equals or exceeds .85 is often used as the cutoff criterion for problematic discriminant validity (cf. guidelines for multicolinearity in regression; Cohen et al., 2003; Tabachnick & Fidell, 2001). When factors overlap to this degree, it may be possible to combine factors to acquire a more parsimonious solution. The goal of such respecification is not to improve the overall fit of the model (e.g., a model with fewer factors entails a smaller number of freely estimated parameters), but ideally the fit of the more parsimonious solution will be similar to the initial model, assuming that overall fit of the initial model was satisfactory, except for the excessive correlation between two factors. Again, revisions of this nature require a clear rationale.

Earlier in this chapter, studies on method effects arising from reverse-worded items were briefly discussed. A paper from this literature provides an applied example of excessive overlap between latent factors. Recall that in the Brown (2003) study, a one-factor model of the PSWQ was hypothesized in which all 16 items were specified to load on a single dimension of pathological worry, with correlated errors to account for method effects from reversed items. Because a two-factor solution ("Worry Engagement," "Absence of Worry"; cf. Fresco et al., 2002) had prevailed in prior EFA and CFA studies of the PSWQ, it was also fit to the data in Brown (2003) to serve as a competing solution to the hypothesized one-factor model. Results of this CFA indicated poor discriminant validity of the Worry Engagement and Absence of Worry latent dimensions; that is, these factors were highly correlated (e.g., r = .87; Brown, 2003). This finding, along with other evidence and considerations (e.g., superior fit of the hypothesized one-factor model, no substantive basis for an "absence of worry" dimension), strongly challenged the acceptability of a two-factor solution.

An alternative to combining two highly correlated factors is to drop one of the factors and its constituent indicators. Although the best remedial approach depends on the specific scientific context, dropping a factor might be favored if one of the factors is defined by only a few indicators or has limited variance, or if substantive and practical considerations support this strategy (e.g., there are clear advantages to developing a briefer questionnaire).

Indicators and Factor Loadings

Another potential source of CFA model misspecification is an incorrect designation of the relationships between indicators and the latent factors. This can occur in the following manners (assuming the correct number of factors was specified): (1) the indicator was specified to load on one factor, but actually has salient loadings on two or more factors; (2) the indicator was specified to load on the wrong factor; and (3) the indicator was specified to load on a factor, but actually has no salient relationship to any factor. Depending on the problem, the remedy will be either to respecify the pattern of relationships between the indicator and the factors or to eliminate the indicator from the model altogether.

Fit diagnostics for these forms of misspecifications are presented in context of the CFA measurement model presented in Figure 5.1. The path diagram in this figure represents the latent structure of the population measurement model; a sample variance–covariance matrix for this model (N = 500) was generated using the Monte Carlo utility in Mplus (see Chapter 10). In this example, a researcher has developed a 12-item questionnaire (items are rated on 0–8 scales) designed to assess young adults' motives to consume alcoholic beverages (cf. Cooper, 1994). The measure was intended to assess three facets of this construct (4 items each): (1) coping motives (items 1–4), (2) social motives (items 5–8), and (3) enhancement motives (items 9–12). All of the items were phrased in the positive direction (e.g., item 5: "Because you feel more self-confident about yourself"), with the exception of items 11 and 12, which are reverse-worded and scored. Although the three-factor model was intended to possess congeneric item sets, the "true" model contains one double-loading item (item 4) and one correlated error resulting from the reverse wording (cf. Marsh, 1996). Figure 5.1 presents the completely standardized parameter estimates and overall fit of this model; LISREL syntax and the input matrix are provided in Table 5.2. As seen in this figure, fit indices were consistent with good model fit; this outcome was supported by standardized residuals below 2.00 and modification indices below 4.00. All unstandardized parameter estimates were statistically significant (ps < .001) and of a magnitude in accord with expectation.

Now consider the various forms of indicator–factor misspecifications previously mentioned. For instance, a poor-fitting model may result from specification of congeneric indicator sets, when in fact some indicators load on more than one factor. In the current example, the researcher is predicting such a model in which four items each load on the three latent factor solution. Specifically, the two-factor solution would provide a model fit identical to the one-factor model if the correlation between Neuroticism and Extraversion was fixed to 1.0. Under this specification, the factors would be identical (i.e., 100% shared variance) and thus the indicators would relate to the factors in the same fashion as in the one-factor solution. Accordingly, the one-factor solution can be viewed as a more constrained version of the two-factor model; that is, in the two-factor model the factor correlation is freely estimated, and thus has one less *df*. This principle generalizes to solutions with larger numbers of factors (e.g., a four-factor solution versus a three-factor solution in which two of the factors from the former model are collapsed).

When models are nested, the χ^2 statistic can be used to statistically compare the fit of the solutions. Used in this fashion, χ^2 is often referred to as the χ^2 difference test (χ^2_{diff}) or the nested χ^2 test. If a model is nested under a parent model, the simple difference in the model χ^2s is also distributed as χ^2 in many circumstances (for exceptions to this guideline, see Chapter 9). For example, in the case of the one- versus two-factor model of Neuroticism and Extraversion, the χ^2 difference test would be calculated as follows:

	df	χ^2
One-factor model	20	373.83
Two-factor model	19	13.23 (see Chapter 4)
χ^2 difference (χ^2_{diff})	1	360.60

Thus, χ^2_{diff} (1) = 360.60. In this example, the χ^2 difference test has 1 *df*, which reflects the difference in model *df*s for the one- and two-factor solutions (20 – 19). Therefore, the critical value for χ^2_{diff} in this example is 3.84 (α = .05, *df* = 1). Because the χ^2_{diff} test value exceeds 3.84 (360.60), it would be concluded that the two-factor model provides a significantly better fit to the data than the one-factor model. It is also important that the two-factor model fit the data well. Use of the χ^2 difference test to compare models is not justified when neither solution provides an acceptable fit to the data.

However, some methodologists would argue that models that differ in regard to the number of latent factors are not nested. This is because the restriction required to make the two-factor model equivalent to the one-factor model (or a three-factor model equivalent to a two-factor model, etc.) entails a fixed parameter that is on the border of admissible parameter space: a factor correlation of 1.0 (i.e., a factor correlation > 1.0 constitutes an out-of-range parameter). In other words, nested models that contain such borderline values (e.g., unit factor correlations, indicator error variances fixed to zero) may not yield proper χ^2 distributions, and thus the χ^2 difference test would also be compromised. Thus, the information criterion indices (e.g., AIC, ECVI; see Eqs. 5.2 and 5.4) could also be considered and reported in instances where the researcher wishes to compare CFA models that differ in the number of factors. These indices are discussed later in this chapter.

Based on the above illustration, the relationship between χ^2_{diff} and modification indices should be more apparent. As discussed in Chapter 4, modification indices represent the predicted decrease in model χ^2 if a fixed or constrained parameter was freely estimated. Accordingly, modification indices (and univariate Lagrange multipliers, cf. EQS) reflect expected χ^2 changes associated with a single *df*. However, nested models often involve solutions that differ by more than a single *df*. For example, a two- and three-factor measurement model of the same data set would differ by two degrees of freedom; that is, the two-factor model contains one factor covariance, the three-factor model contains three factor covariances. Measurement invariance evaluation (Chapter 7) typically entails simultaneously placing constraints on multiple parameters (e.g., constraining factor loadings to equality in two or more groups). Such models are also nested, but differ by more than a single *df*. Moreover, modification indices differ from χ^2_{diff} test values in that they represent an *estimate* of how much model χ^2 will decrease after freeing a fixed or constrained parameter. Quite often, the *actual* difference in model χ^2 (reflected by χ^2_{diff}) produced by a single *df* model modification differs from the estimate of χ^2 change provided by the modification index (this is also true for expected parameter change values; see Chapter 4).

Although applied researchers often compare CFA measurement models that differ in number of latent factors, it is important that a strong conceptual rationale exist for doing so. Occasionally, such analyses are "straw man" comparisons, where the models specified as competing solutions to the hypothesized model have dubious substantive bases and little likelihood of providing an equivalent or superior fit to the data; for example, in the psychometric evaluation of a questionnaire, comparing a three-factor model to a one-factor model, when in fact the measure was designed to be multifactorial and prior EFA research has supported this structure. Whereas the one- versus two-factor models of Neuroticism and Extraversion were presented in this chapter to discuss some of the concepts and issues of model comparison, this criticism would apply; that is, there was no basis for the one-factor model, in view of compelling theory and scien-

TABLE 5.1. Standardized Residuals and Modification for a One-Factor CFA Model of Indicators of Neuroticism and Extraversion

Standardized Residuals

	N1	N2	N3	N4	E1	E2	E3	E4
N1	- -							
N2	2.5394	- -						
N3	-1.1329	-0.0994	- -					
N4	1.2744	0.3203	2.5275	- -				
E1	1.5849	2.5164	0.5238	2.7600	- -			
E2	1.5119	2.0128	1.3017	3.3384	9.7512	- -		
E3	2.4183	1.7140	1.4378	2.1656	8.9485	9.3798	- -	
E4	1.6615	1.8858	0.5442	2.9684	7.1715	8.3930	7.8846	- -

Modification Indices for THETA-DELTA

	N1	N2	N3	N4	E1	E2	E3	E4
N1	- -							
N2	6.4488	- -						
N3	1.2834	0.0099	- -					
N4	1.6243	0.1026	6.3884	- -				
E1	2.5118	6.3322	0.2744	7.6176	- -			
E2	2.2858	4.0513	1.6945	11.1448	95.0860	- -		
E3	5.8479	2.9376	2.0673	4.6898	80.0762	87.9802	- -	
E4	2.7604	3.5562	0.2962	8.8113	51.4310	70.4417	62.1671	- -

Note. Results obtained from LISREL 8.72.

outcomes signify distinct factors versus method effects (or *minor factors*) is not always clear.

The nested model comparison methodology is often used in the applied literature to statistically compare the fit of CFA models that differ in terms of the number of latent factors (e.g., does a two-factor model of Neuroticism and Extraversion provide a better fit to the data than a one-factor model?). Recall that a *nested model* contains a subset of the freed parameters of another solution (Chapters 3 and 4). Consider the following one-factor models involving the same set of 5 indicators (number of elements of the input variance–covariance matrix = 5(6)/2 = 15 = 5 variances, 10 covariances). *Model N* (the nested model) is a simple one-factor solution with no correlated measurement errors; thus, it consists of 10 freed parameters (5 factor loadings, 5 indicator errors, factor variance is fixed to 1.0 to define latent variable metric), and the model has 5 df (15 – 10). *Model P* (the parent model) is identical to Model N, with the exception that a correlated error is specified for the fourth and fifth indicators; thus, it consists of 11 freely estimated parameters (5 factor loadings, 5 indicator errors, 1 correlated error), and this model's df = 4 (15 – 11). In this scenario, Model N is nested under Model P; if a path is dropped from Model P—the correlated residual for indicators 4 and 5—Model N is formed. In other words, Model N contains a subset of the freed parameters of the parent model, Model P; a nested model will possess a larger number of dfs than the parent model, in this case the df difference is 1, that is, 5 – 4. In Model N, the correlations among all indicator errors are fixed to zero. In Model P, this is not the case because the correlation between the residuals of two indicators is freely estimated.

Models that differ in the number of latent factors are considered nested models. For example, consider the one- and two-factor solutions for the 8 indicators of Neuroticism and Extraversion (b = 36 = 8 variances and 28 covariances). The two-factor solution discussed in Chapter 4 contains 17 freely estimated parameters: 6 loadings, 8 indicator errors, 2 factor variances, 1 factor covariance (the loadings of N1 and E1 were fixed as marker indicators). Thus, this two-factor model has df = 19 (36 – 17). In contrast, the one-factor solution presented above contains 16 freed parameters: 7 loadings, 8 errors, 1 factor variance (the loading of N1 was fixed as the marker indicator). The one-factor solution has df = 20 (36 – 16). Although structurally more discrepant than the Model N versus Model P example described previously, the one-factor model could be construed as nested under the two-factor solution, again with a difference of df = 1 (i.e., 20 – 19). This single df difference relates to factor correlation in the two the general assessment modality (e.g., questionnaires, behavioral observation ratings, clinical interview ratings). Or more specifically, these effects may be due to similarly or reverse-worded assessment items, or other sources such as items with differential proneness to response set, acquiescence, or social desirability.[1] A comprehensive review of the potential sources of method effects is provided in Podsakoff, MacKenzie, Lee, and Podsakoff (2003).

The influence of method effects has been illustrated in factor analyses of the Self-Esteem Scale (SES; Rosenberg, 1965) (e.g., Marsh, 1996; Tomás & Oliver, 1999; Wang, Siegal, Falck, & Carlson, 2001) and the Penn State Worry Questionnaire (PSWQ; Meyer, Miller, Metzger, & Borkovec, 1990) (e.g., Brown, 2003; Fresco, Heimberg, Mennin, & Turk, 2002). Specifically, this research has shown the impact of method effects in questionnaires comprised of some mixture of positively and negatively worded items (the SES contains 4 positively worded items, e.g., "I feel good about myself," and 3 negatively worded items, e.g., "At times I think I am no good at all"; the PSWQ contains 11 items worded in the symptomatic direction, e.g., "I worry all the time," and 5 items worded in the nonsymptomatic direction, e.g., "I never worry about anything"). In other words, the differential covariance among these items is not based on the influence of distinct, substantively important latent dimensions. Rather, this covariation reflects an artifact of response styles associated with the wording of the items (cf. Marsh, 1996).

Nevertheless, studies that conducted EFAs with these measures routinely reported two-factor solutions, with one factor comprised of positively worded items (SES: "Positive Self-Evaluation;" PSWQ: "Absence of Worry") and the second factor comprised of negatively worded items (SES: "Negative Self-Evaluation"; PSWQ: "Worry Engagement"). However, subsequent CFA studies challenged the conceptual utility of these two-factor models (e.g., Brown, 2003; Hazlett-Stevens, Ullman, & Craske, 2004; Marsh, 1996). For example, because the PSWQ was designed to measure the trait of pathological worry, what is the practical and conceptual importance of a dimension of "absence of worry" (Brown, 2003)? In both lines of research, the CFA studies (e.g., Brown, 2003; Marsh, 1996) demonstrated the substantive (i.e., interpretability) and empirical (i.e., goodness of fit) superiority of single-factor solutions where the additional covariance stemming from the directionality of item wording was accounted for by correlated measurement errors (for a detailed discussion of this approach, see Chapter 6). This also highlights the importance of keeping substantive issues firmly in mind when formulating and interpreting EFA and CFA solutions. In the above examples, two-factor EFA solutions for SES and PSWQ items are apt to provide a better fit (in terms of χ^2, RMSEA, etc.) than a one-factor model. Although the viability of a one-factor model with correlated errors could not be explored within the EFA framework, the acceptability of these two-factor solutions could be challenged on substantive grounds, despite their superior fit.

In the case of a typical measurement model of congeneric indicator sets (a model in which there are no double-loading items and no correlated errors), a CFA solution with too few factors will fail to adequately reproduce the observed relationships among several indicators. For instance, consider the scenario where the two-factor model of Neuroticism and Extraversion in Chapter 4 is specified as a one-factor solution with no correlated indicator errors (ML estimation). The overall fit of this solution is poor, $\chi^2(20) = 373.83$, $p < .001$, SRMR = .187, RMSEA = .306 (90% CI = .283 to .330; CFit $p < .001$), TLI = .71, CFI = .79. As seen in Table 5.1, both the standardized residuals and modification indices indicate that, as a consequence of forcing the four Extraversion indicators (E1–E4) to load on the same latent variable as the four indicators of Neuroticism (N1–N4), the parameter estimates of the solution markedly underestimate the observed relationships among the E1–E4 measures (see Chapter 4 for guidelines on interpreting modification indices and standardized residuals). Specifically, the range of standardized residuals for the Extraversion indicators range from 7.17 to 9.75, and modification indices range from 51.43 to 91.09. It is noteworthy that in the case of a one-factor solution, modification indices can appear only in sections of the results that pertain to indicator measurement errors (e.g., the Theta-Delta matrix in LISREL; see Table 5.1). Although the "true" model in this instance is a two-factor solution (see Figure 4.2, Chapter 4), the fact that fit diagnostics appear only in this fashion might lead novice CFA researchers to conclude that correlated measurement errors are required. Modification indices can point to problems with the model that are not the real source of mis-fit. Again, this underscores the importance of an explicit substantive basis (both conceptual and empirical) for model (re)specification; for example, specifying a model with correlated errors among the Extraversion indicators is not well founded in relation to a model entailing two distinct factors. In this example, the pattern of relationships in the input matrix (see Figure 4.1, Chapter 4), in addition to the aggregation of standardized residuals and modification indices associated with a set of indicators (E1–E4; see Table 5.1), would provide clear empirical evidence against a simple one-factor solution. However, the determination of whether such evaluate the acceptability of the model are not satisfied; that is, the model does not fit well on the whole, does not reproduce some indicator relationships well, or does not produce uniformly interpretable parameter estimates (see Chapter 4). Based on fit diagnostic information (e.g., modification indices) and substantive justification, the model is revised and fit to the data again in the hope of improving its goodness of fit. The sources of CFA model misspecification, and the methods of detecting and rectifying those sources, are discussed in the next section of this chapter.

In addition, respecification is often conducted to improve the parsimony and interpretability of the CFA model. Rarely do these forms of respecification improve the fit of the solution; in fact, they may worsen overall fit to some degree. For example, the results of an initial CFA may indicate that some factors have poor discriminant validity—that is, two factors are so highly correlated that the notion that they represent distinct constructs is untenable. Based on this outcome, the model may be respecified by collapsing the highly overlapping factors; that is, the indicators that loaded on separate, overlapping factors are respecified to load on a single factor. Although this respecification may foster the parsimony and interpretability of the measurement model, it will lead to some decrease in goodness of fit relative to the more complex initial solution.

Two other types of model respecification frequently used to improve parsimony are multiple-groups solutions and higher-order factor models. These respecifications are conducted after an initial CFA model has been found to fit the data well. For instance, equality constraints are placed on parameters in multiple-groups CFA solutions (e.g., factor loadings) to determine the equivalence of a measurement model across groups (e.g., do test items show the same relationships to the underlying construct of cognitive ability in men and women?). With the possible exception of parsimony fit indices (e.g., RMSEA, TLI), these constraints will not improve the fit of the model as compared with a baseline solution where the parameters are freely estimated in all groups. In higher-order CFA models, the goal is to reproduce the correlations among the factors of an initial CFA solution with a more parsimonious higher-order factor structure; for example, can the six correlations among the four factors of an initial CFA model be reproduced by a single higher-order factor? As with the previous examples, this respecification cannot improve model fit because the number of parameters in the higher-order factor structure is less than the number of freely estimated factor correlations in the initial CFA model. Multiple-groups solutions and higher-order factor models are discussed in Chapter 7 and Chapter 8, respectively.

SOURCES OF POOR-FITTING CFA SOLUTIONS

In a CFA model, the main potential sources of misspecification are the number of factors (too few or too many), the indicators (e.g., selection of indicators, patterning of indicator–factor loadings), and the error theory (e.g., uncorrelated vs. correlated measurement errors). As discussed in Chapter 4, a misspecified CFA solution may show itself in several aspects of the results: (1) overall goodness-of-fit indices that fall below accepted thresholds (e.g., CFI, TLI < .95); (2) large standardized residuals or modification indices; and (3) unexpectedly large or small parameter estimates or "Heywood cases," which are estimates with out-of-range values. Standardized residuals and modification indices are often useful for determining the particular sources of strain in the solution. However, these statistics are most apt to be helpful when the solution contains minor misspecifications. When the initial model is grossly misspecified, specification searches are not nearly as likely to be successful (MacCallum, 1986).

Number of Factors

In practice, misspecifications resulting from an improper number of factors should occur rarely. When this occurs, it is likely that the researcher has moved into the CFA framework prematurely. CFA hinges on a strong conceptual and empirical basis. Thus, in addition to a compelling substantive justification, CFA model specification is usually supported by prior (but less restrictive) exploratory analyses (i.e., EFA) that have established the appropriate number of factors, and pattern of indicator–factor relationships. Accordingly, gross misspecifications (e.g., specifying too many or too few factors) should be unlikely when the proper groundwork for CFA has been conducted.

However, there are some instances where EFA has the potential to provide misleading information regarding the appropriate number of factors in CFA. This is particularly evident in scenarios where the relationships among indicators are better accounted for by correlated errors than separate factors. A limitation of EFA is that its identification restrictions preclude the specification of correlated indicator errors (see Chapters 2 and 3). Thus, the results of EFA may suggest additional factors when in fact the relationships among some indicators are better explained by correlated errors from *method effects*. A method effect exists when some of the differential covariance among items is due to the measurement approach rather than the substantive latent factors. Method effects may stem from model fit was defined by the following criteria: RMSEA ($\leq .06$, 90% CI $\leq .06$, CFit *ns*), SRMR ($\leq .08$), CFI ($\geq .95$), and TLI ($\geq .95$). Multiple indices were used because they provide different information about model fit (i.e., absolute fit, fit adjusting for model parsimony, fit relative to a null model); used together, these indices provide a more conservative and reliable evaluation of the solution.

Each of the overall goodness-of-fit indices suggested that the two-factor model fit the data well, $\chi^2(19) = 13.23$, $p = .83$, SRMR = .019, RMSEA = 0.00 (90% CI = 0.00 – .018; CFit = .99), TLI = 1.007, CFI = 1.00. Inspection of standardized residuals and modification indices indicated no localized points of ill fit in the solution (e.g., largest modification index = 3.49, largest standardized residual = 1.87). Unstandardized and completely standardized parameter estimates from this solution are presented in Figure 4.2 (standard errors of the estimates are provided in Table 4.4). All freely estimated unstandardized parameters were statistically significant (*ps* < .001). Factor loading estimates revealed that the indicators were strongly related to their purported latent factors (range of R^2s = .49–.78), consistent with the position that the NEO scales are reliable indicators of the constructs of neuroticism and extraversion. Moreover, estimates from the two-factor solution indicate a moderate relationship between the dimensions of Neuroticism and Extraaversion (–.435), in accord with previous evidence and theory.

<Discussion of other implications, limitations, and future directions would follow>

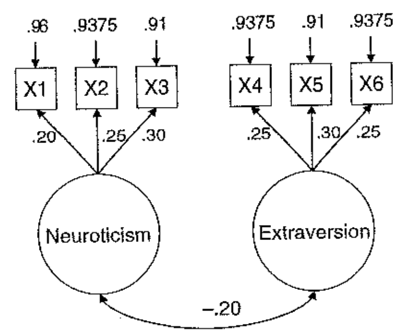
5

CFA Model Revision and Comparison

In Chapter 4, many of the procedures and issues associated with model specification were introduced in the context of a good-fitting, two-factor CFA solution of Neuroticism and Extraversion. In this chapter, these concepts are further illustrated and extended, using initially poor-fitting solutions. Although poor-fitting models are frequently encountered in applied research, SEM sourcebooks rarely deal with this topic in detail because of the numerous potential sources of ill-fit, and the fact that proper model specification and respecification hinges directly on the substantive context of the analysis. Thus, although some examples are provided, readers must adapt these general guidelines and principles to the specific aspects of their own data sets and models. This chapter discusses the fact that model respecification can also be carried out to improve the parsimony and interpretability of a CFA solution. The sources of improper solutions, and ways to deal with them, are also discussed. In addition, the technique of EFA within the CFA framework (E/CFA) is presented as an intermediate step between EFA and CFA. This methodology allows the researcher to explore measurement structures more fully to develop more realistic (i.e., better-fitting) CFA solutions. The chapter concludes with a discussion of equivalent CFA models, an issue that is germane to virtually all measurement models but is largely unrecognized in applied research.

GOALS OF MODEL RESPECIFICATION

Often a CFA model will need to be revised. The most common reason for respecification is to improve the fit of the model. In this case, the results of an initial CFA indicate that one or more of the three major criteria used to



In fact, this solution provides a perfect fit to the data, $\chi^2(8) = 0.00$; for example, multiplying the factor loadings of X1 and X2 perfectly reproduces their observed relationship ($.20 \times .25 = .05$), multiplying the factor loading of X1 and X5 with the factor correlation perfectly reproduces the observed relationship between X1 and X5 ($.20 \times .30 \times -.20 = -.012$). However, goodness of fit of the two-factor solution is not determined by the *absolute magnitude* of the sample correlations, but by whether the *differential magnitude* of the correlations can be reproduced by the specified model and its resulting parameter estimates. Accordingly, because the X1–X3 indicators are more strongly intercorrelated with each other than they are with the X4–X6 indicators (and vice versa), the two-factor model specification fits the data well. Nevertheless, the size of the resulting parameter estimates would lead the sensible researcher to reject the two-factor model and the questionnaire on which it is based. For example, the latent factor of Neuroticism accounts for only 4% of the variance in the X1 indicator ($.20^2 = .04$); that is, the greatest part of its variance (96%) is not explained by the latent variable. Indeed, the largest percentage of variance explained in the items is only 9% (X3 and X5). Thus, the questionnaire items are very weakly related to the latent factors and should not be considered to be reasonable indicators of their purported constructs.

Appendix 4.3

Example Report of the Two-Factor CFA Model of Neuroticism and Extraversion

Based on prior evidence and theory bearing on the Big Five model of personality, a two-factor model was specified in which anxiety (N1), hostility (N2), depression (N3), and self-consciousness (N4) loaded onto the latent variable of Neuroticism, and in which warmth (E1), gregariousness (E2), assertiveness (E3), and positive emotions (E4) loaded onto the latent variable of Extraversion. The indicators were subscales of the NEO Personality Inventory and had a range of scores from 0 to 32, with higher scores reflecting higher levels of the personality dimension. Figure 4.1 depicts the complete specification of the two-factor model. Anxiety (N1) and warmth (E1) were used as marker indicators for Neuroticism and Extraversion, respectively. The measurement model contained no double-loading indicators and all measurement error was presumed to be uncorrelated. The latent factors of Neuroticism and Extraversion were permitted to be correlated based on prior evidence of a moderate inverse relationship between these dimensions. Accordingly, the model was overidentified with 19 *df*.

As noted in the Method section, the NEO was administered to 250 college undergraduates who participated in the study for course credit (see Method section for a description of sample demographics). All 250 cases had complete NEO data. Prior to the CFA analysis, the data were evaluated for univariate and multivariate outliers by examining leverage indices for each participant. An outlier was defined as a leverage score that was five times greater than the sample average leverage value. No univariate or multivariate outliers were detected. Normality of the indicators was examined using PRELIS 2.30 (Jöreskog & Sörbom, 1996b). The standardized skewness score was 1.17 and the standardized kurtosis score was 1.32 ($ps > .05$). The joint test of non-normality in terms of skewness and kurtosis was not significant, $\chi^2 = 8.35$, $p = .38$.

Thus, the sample variance–covariance matrix was analyzed using LISREL 8.72, and a maximum likelihood minimization function (sample correlations and *SD*s are provided in Figure 4.1). Goodness of fit was evaluated using the standardized root mean square residual (SRMR), root mean square error of approximation (RMSEA) and its 90% confidence interval (90% CI) and test of close fit (CFit), comparative fit index (CFI), and the Tucker–Lewis index (TLI). Guided by suggestions provided in Hu and Bentler (1999), acceptable be preferable to make the data available by request or possibly by downloading from the author's or journal's website.

Finally, it should be emphasized that the suggestions provided in Table 4.6 are most germane to a measurement model conducted in a single group and thus must be adapted on the basis of the nature of the particular CFA study. For example, many CFA investigations entail the comparison of a target model to a substantively viable competing solution (see Chapter 5). In such studies, the recommendations listed in Table 4.6 should be extended to the alternative models as well (e.g., also provide conceptual/empirical justification for the competing models). In multiple-group CFA studies (see Chapter 7), it is important to evaluate the CFA solutions separately in each group before conducting the simultaneous analysis. In addition, the multiple-groups analysis typically entails invariance evaluation of the CFA parameters (e.g., are the factor loadings invariant, consistent with the notion that the indicators measure the latent construct in comparable ways in all groups?). Thus, although all of the steps listed in Table 4.6 are relevant, the report of a multiple-groups CFA study is considerably more extensive than the sample write-up provided in Appendix 4.3 (e.g., data normality screening and model fit estimation/evaluation in each group, nested invariance evaluation; see Chapter 7).

SUMMARY

The fundamental concepts and procedures of CFA were illustrated using a full example. As can be seen in this chapter, the proper conduct of CFA requires a series of steps and decisions including the specification of the measurement model (based on prior evidence and theory), selection of a statistical estimator appropriate for the type and distributional properties of the data (e.g., ML), choice of a latent variable software program (e.g., EQS, LISREL), evaluation of the acceptability of the model (e.g., overall goodness of fit, focal areas of strain in the solution, interpretability/strength of parameter estimates), and the interpretation and presentation of results. Although this material was presented in the context of a well-specified measurement model, some of the complications and issues often encountered in applied CFA research were introduced (e.g., potential sources of ill fit, Heywood cases, significance testing of parameter estimates). These issues are considered at much greater length in the next chapter, which focuses on the respecification and comparison of CFA models.

NOTES

1. Although in the case of the Figure 4.1 model, specification of a direct effect between Neuroticism and Extraversion would produce the same fit as a solution that simply allows these factors to be intercorrelated. This is because the potential structural component of the model (e.g., Neuroticism → Extraversion, or Extraversion → Neuroticism) is just-identified.

2. However, in EQS, the term *standardized residual* is used differently. In EQS, a standardized residual reflects the difference between the observed correlation and the model-implied correlation; for example, for the N1 and N2 relationship in the Figure 4.1 model, the EQS standardized residual is .016 (i.e., .767 – .751).

3. Like Lagrange multipliers, Wald tests can be used as multivariate statistics that estimate the change in model χ^2 if *sets* of freed parameters are fixed.

4. Although completely standardized loadings > 1.0 are generally considered to be Heywood cases, Jöreskog (1999) has demonstrated instances where such estimates are valid (i.e., models that contain double-loading indicators).

5. One-tailed (directional) tests are appropriate for parameter estimates involving variances (i.e., indicator error variances, factor variances) because these parameters cannot have values below zero (z_{crit} = 1.645, α = .05, one-tailed).

6. For indicators that load on more than one factor, factor loadings should be interpreted as partial regression coefficients; for example, a given factor loading for a double-loading indicator would be interpreted as how much the indicator is predicted to change given a unit increase in one factor, holding the other factor constant.

7. The question often arises as to which reliability estimate should be selected for the error variance constraint in situations where the measure in question has an extensive psychometric history or the quality of the extant psychometric evidence is poor. Although qualitative considerations are important (e.g., selection guided by the quality and generalizability of psychometric studies), it is often useful to conduct a sensitivity analysis in which the stability of the results is examined using a range of viable reliability estimates.

TABLE 4.6. *(cont.)*

- Parameter estimates
  - —Provide all parameter estimates (e.g., factor loadings, error variances, factor variances), including any nonsignificant estimates
  - —Consider the clinical as well as the statistical significance of the parameter estimates (e.g., are all indicators meaningfully related to the factors?)
  - —Ideally, include the standard errors or confidence intervals of the parameter estimates
- If necessary (e.g., suitability of *N* could be questioned), report steps taken to verify the power and precision of the model estimates (e.g., Monte Carlo evaluation using the model estimates as population values[c])

Substantive Conclusions

- Discuss CFA results in regard to their substantive implications, directions for future research, and so on.
- Interpret the findings in context of study limitations (e.g., range and properties of the indicators and sample) and other important considerations (e.g., equivalent CFA models[d])

[a]see Chapter 7; [b]see Chapter 9; [c]see Chapter 10; [d]see Chapter 5.

McDonald & Ho, 2002). These data could also be readily presented in a table or figure note, or in the body of the table or figure itself when only unstandardized estimates are provided.

It should be noted that there is no "gold standard" for how a path diagram should be prepared. Although some constants exist (e.g., representing observed and latent variables by rectangles and circles, respectively; depicting a direct effect by a unidirectional arrow), the reader will encounter many variations in the applied and quantitative literature; for instance, an indicator error variance may be represented by "e," "θ," "δ," "ε," a circle, a double-headed curved arrow, or some combination. No approach is considered to be more correct than another. One particular method is used throughout this book, but the reader is encouraged to peruse other sources to decide which approach best suits his or her own tastes and purposes.

Another consideration is whether to present an unstandardized solution, completely standardized solution, or both (also, as noted in the previous section, standardized estimates may be relevant for some parameters such as in MIMIC models). In applied CFA research, the convention has been to report completely standardized parameters. This may be due in part to the fact that CFA is frequently employed after EFA in the psychometric analysis of test instruments (e.g., multiple-item questionnaires). As noted in Chapter 2, the tradition of EFA is to standardize both the observed and latent variables. Often, neither the observed nor latent variables structure is standardized in the CFA analysis (e.g., a variance–covariance matrix is inputted, a marker indicator is specified). Although the completely standardized CFA solution can be informative (e.g., a completely standardized error indicates the proportion of variance in the indicator that is not accounted for by the latent factor), the unstandardized solution (or both) may be preferred in some instances such as in measurement invariance evaluations where constraints are placed on the unstandardized parameters (Chapter 7), in construct validity studies where the indicators are composite measures with readily interpretable metrics, and in analyses using item parcels (Chapter 9) where information regarding the relative magnitudes of the relationships to the factors has little substantive importance or does not convey information about the original items. As discussed in Chapter 3, there are some CFA and SEM scenarios where exclusive reliance on completely standardized solutions can result in misleading or erroneous conclusions (Bollen, 1989; Willett et al., 1998).

If possible, the sample input data used in the CFA should be published in the research report. In instances where this is not feasible (e.g., a large set of indicators are used), it is helpful for authors to make these data available upon request (or post the data on a website). Inclusion of these data provides a wealth of information (e.g., magnitudes and patterns of relationships among variables) and allows the reader to replicate the study's models and to explore possible conceptually viable alternative models (equivalent or better fitting; see Chapter 5). In general, the sample correlation matrix (accompanied with *SD*s and *M*s, if applicable) should be provided, rather than a variance–covariance matrix. This is because the reader will be able to analyze a variance–covariance matrix that contains less rounding error than the typical variance–covariance matrix published in research reports by creating it directly from the sample correlations and *SD*s. Of course, there are some situations where the CFA analysis cannot be reproduced from a published table of data. Whereas tabled data work fine for ML analysis of data sets that contain no missing data (or where pairwise or listwise deletion was used to manage missing data), this is not the case for analyses that are conducted on raw data (e.g., as in analyses that use direct ML for missing data; see Chapter 9) or that require companion matrices (e.g., asymptotic covariance as well as tetrachoric correlation matrices in WLS-estimated models; see Chapter 9). In these cases, it may

Correlations involving the general well-being factor (GWB) reflect the relationship of the GENHLTH indicator with other variables in the model, adjusting for measurement error; for example, the correlation of GENHLTH with Physical Functioning is .642. The factor loading of age on the Age "factor" is 1.00, reflecting a perfect relationship between the observed measure and its underlying "true" score.

REPORTING A CFA STUDY

Good-fitting CFA solutions are presented in this chapter to illustrate the specification and interpretation of latent variable measurement models. In addition to understanding how to conduct and interpret these analyses properly, applied researchers must be aware of what information should be presented when reporting the results of a CFA study. Indeed, although many excellent guides have been published for reporting SEM and CFA results (e.g., Hoyle & Panter, 1995; McDonald & Ho, 2002; Raykov, Tomer, & Nesselroade, 1991), recent surveys have indicated that applied research articles continue to often omit key aspects of the analyses such as the type of input matrix used, the identifiability and exact specification of the model, and the resulting parameter estimates (MacCallum & Austin, 2000; McDonald & Ho, 2002). Recommended information to include in a CFA research report is listed in Table 4.6. Appendix 4.3 provides a sample write-up of some of this suggested information using the example of the two-factor CFA model of Neuroticism and Extraversion.

Although Table 4.6 and Appendix 4.3 are fairly self-explanatory, a few elaborations are warranted. The parameter estimates from the two-factor model used in this example lent themselves to presentation in a path diagram (Figure 4.2), but many applied CFA models are too complex to be presented in this fashion (e.g., models often contain a large number of indicators and latent factors). Thus, applied CFA findings are frequently presented in tabular formats; specifically, a p (indicator) by m (factor) matrix of factor loadings. While a tabular approach may be preferred for large CFA models (cf. McDonald & Ho, 2002), the researcher must be sure to provide the remaining parameter estimates (e.g., factor and error correlations) in the text, a table note, or a companion figure. Although methodologists have underscored the importance of providing standard errors (and/or confidence intervals) of parameter estimates, this information is rarely reported in CFA research (MacCallum & Austin, 2000;

TABLE 4.6. Information to Report in a CFA Study

Model Specification

- Conceptual/empirical justification for the hypothesized model
- Complete description of the parameter specification of the model
  - —List the indicators for each factor
  - —Indicate how the metric of the factors was defined (e.g, specify which observed variables were used as marker indicators)
  - —Describe all freely estimated, fixed, and constrained parameters (e.g., factor loadings and cross-loadings, random and correlated indicator errors, factor correlations, intercepts and factor means[a])
- Demonstrate that the model is identified (e.g., positive model *df*, scaling of latent variables, absence of empirical underidentification)

Input Data

- Description of sample characteristics, sample size, and sampling method
- Description of the type of data used (e.g., nominal, interval; scale range of indicators)
- Tests of estimator assumptions (e.g., multivariate normality of input indicators)
- Extent and nature of missing data, and the method of missing data management (e.g., direct ML, multiple imputation[b])
- Provide sample correlation matrix and indicator *SDs* (and means, if applicable[a]), or make such data available on request

Model Estimation

- Indicate the software and version used (e.g., LISREL 8.72)
- Indicate the type of data/matrices analyzed (e.g., variance–covariance, tetrachoric correlations/asymptotic covariances[b])
- Indicate the estimator used (e.g., ML, weighted least squares[b]; as justified by properties of the input data)

Model Evaluation

- Overall goodness-of-fit
  - —Report model χ^2 along with its *df* and *p* value
  - —Report multiple fit indices (e.g., SRMR, RMSEA, CFI) and indicate cutoffs used (e.g., RMSEA $\leq$ .06); provide confidence intervals, if applicable (e.g., RMSEA)
- Localized areas of ill fit
  - —Report strategies used to assess for focal strains in the solution (e.g., modification indices/Lagrange multipliers, standardized residuals, Wald tests, EPC values)
  - —Report absence of areas of ill fit (e.g., largest modification index) or indicate the areas of strain in the model (e.g., modification index, EPC value)
- If model is respecified, provide a compelling substantive rationale for the added or removed parameters and clearly document (improvement in) fit of the modified models

(con