Byrne, B. M., 2006: Structural Equation Modeling with EQS. Basic Concepts, Applications and Programming (2nd edition). Mahwah: Erlbaum, S. 250-260.

8

Application 6: Testing for the Invariance of a Causal Structure

In chapter 4, several problematic aspects of post hoc model-fitting in SEM were highlighted. Indeed, so common is this practice and so frequently is it conducted with little regard for the substantive meaningfulness of the respecified models that concerned researchers have provided various means by which such models can be tested more stringently (see, e.g., Anderson & Gerbing, 1988; Cudeck & Henly, 1991; MacCallum, 1995; and MacCallum et al., 1992, 1993). It was noted in chapter 4 that one approach to addressing problems associated with post hoc model-fitting is to apply some mode of cross-validation analysis; this is the focus of the present chapter. The application demonstrated herein is a follow-up to the analyses of chapter 6 in which the validity of a causal structure describing determinants of burnout for a calibration sample of secondary school teachers was tested. However, before a walk-through of this cross-validation procedure, it is important to review some of the related issues.

CROSS-VALIDATION IN SEM

Typically in applications of SEM, the researcher tests a hypothesized model and then, from an assessment of various goodness-of-fit criteria, concludes that a statistically better-fitting model could be attained by respecifying the model such that particular parameters previously constrained to zero are freely estimated (Breckler,

1990; MacCallum et al., 1992, 1993; and MacCallum, Roznowski, Mar, & Reith, 1994). Possibly as a consequence of considerable criticism of SEM procedures in the past (e.g., Biddle & Marlin, 1987; Breckler, 1990; and Cliff, 1983), most researchers who proceed with this respecification process are now generally familiar with the issues. In particular, they are cognizant of the exploratory nature of these follow-up procedures as well as the fact that additionally specified parameters in the model must be theoretically substantiated.

Indeed, the pros and cons of post hoc model-fitting have been rigorously debated in the literature. Although some have severely criticized the practice (e.g., Cliff, 1983; and Cudeck & Browne, 1983), others have argued that as long as researchers are fully cognizant of the exploratory nature of their analyses, the process can be substantively meaningful because both practical and statistical significance can be considered (Byrne et al., 1989; and Tanaka & Huba, 1984). However, Jöreskog (1993, p. 298) is clear in stating that "If the model is rejected by the data, the problem is to determine what is wrong with the model and how the model should be modified to fit the data better." The purists would argue that once a hypothesized model is rejected, it is the end of the story. More realistically, however, other researchers in this area recognize the obvious impracticality in the termination of all subsequent model analyses. Clearly, in the interest of future research, it behooves the investigator to probe deeper into the question of why the model is misfitting (Tanaka, 1993). As a consequence of the concerted efforts of statistical experts in addressing this issue, there are now several different approaches that can be used to increase the soundness of findings derived from these post hoc analyses.

Undoubtedly, post hoc model-fitting in the analysis of covariance structures is problematic. With multiple model specifications, there is a risk of capitalization on chance factors because model modification may be driven by characteristics of the particular sample on which the model was tested (e.g., sample size, sample heterogeneity) (MacCallum et al., 1992). As a consequence of this sequential testing procedure, there is increased risk of making either a Type I or Type II error and, at this point, there is no direct way to adjust for the probability of such error. Because hypothesized covariance structure models represent only approximations of reality and thus are not expected to fit real-world phenomena exactly (Cudeck & Browne, 1983; and MacCallum et al., 1992), most research applications are likely to require the specification of alternative models in the quest for one that fits the data well (Anderson & Gerbing, 1988; and MacCallum, 1986). Indeed, this aspect of covariance structure modeling represents a serious limitation; to date, several alternative strategies for model testing have been proposed (see, e.g., Anderson & Gerbing, 1988; Cudeck & Henly, 1991; MacCallum, 1995; and MacCallum et al., 1992, 1993).

One approach to addressing problems associated with post hoc model fitting is to employ a cross-validation strategy, whereby the final model derived from the post hoc analyses is tested on a second (or more) independent sample(s) from the same population. Barring the availability of separate data samples, albeit a sufficiently large sample, the researcher may want to randomly split the data into two (or more) parts, thereby making it possible to cross-validate the findings (Cudeck & Browne, 1983). As such, Sample A serves as the calibration sample on which the initially hypothesized model is tested as well as any post hoc analyses conducted in the process of attaining a well-fitting model. Once this final model is determined, the validity of its structure can then be tested based on Sample B (the validation sample). In other words, the final best-fitting model for the calibration sample becomes the hypothesized model under test for the validation sample.

There are several ways to test the similarity of model structure (e.g., MacCallum et al., 1994). For example, Cudeck and Browne (1983) suggested the computation of a Cross-Validation Index (CVI) that measures the distance between the restricted (i.e., model-imposed) variance–covariance matrix for the calibration sample and the unrestricted variance–covariance matrix for the validation sample. Because the estimated predictive validity of the model is gauged by the smallness of the CVI value, evaluation is facilitated by their comparison based on a series of alternative models. It is important to note, however, the CVI estimate reflects overall discrepancy between "the actual population covariance matrix, Σ, and the estimated population covariance matrix reconstructed from the parameter estimates obtained from fitting the model to the sample" (MacCallum et al., 1994, p. 4). More specifically, this global index of discrepancy represents combined effects arising from the discrepancy of approximation (e.g., nonlinear influences among variables) and the discrepancy of estimation (e.g., representative sample; sample size). (For more discussion of these discrepancy aspects, see Browne & Cudeck, 1989; Cudeck & Henly, 1991; and MacCallum et al., 1994.)

In this chapter, we examine another approach to cross-validation. Specifically, we use an invariance-testing strategy to test for the replicability of both the measurement and structural models across groups. Although there are numerous approaches to cross-validation in SEM depending on the focus of the study (Anderson & Gerbing, 1988; Browne & Cudeck, 1989; and Cudeck & Browne, 1983), the application described herein is straightforward in addressing the question of whether a model that has been respecified in one sample replicates over a second independent sample from the same population. (For another approach in addition to the one demonstrated here, see Byrne & Baron, 1994.)

TESTING FOR INVARIANCE ACROSS CALIBRATION/VALIDATION SAMPLES

The present example comes from the same study briefly described in chapter 6 (Byrne, 1994c). The intent of this original study was to (a) validate a causal structure involving the impact of organizational and personality factors on three facets of burnout for elementary, intermediate, and secondary school teachers; (b) cross-validate this model across a second independent sample within each teaching panel; and (c) test for the invariance of common structural paths across teaching panels. In chapter 6, we tested for the validity of this causal structure for the "calibration" sample of secondary school teachers ($n = 716$); following several respecifications of the model, a final best-fitting model was determined that was deemed valid from the perspective of substantive meaningfulness. (For an in-depth examination of invariance-testing procedures within and between the three teacher groups, see Byrne, 1994b.) The task in this chapter is to cross-validate this final model of burnout structure across a second independent group of high school teachers, which we'll call the "validation group" ($n = 714$).

Although the present example of cross-validation is based on a full structural equation model, the practice is in no way limited to such applications. Indeed, cross-validation is equally as important for CFA models, and examples of such applications are found in various disciplines: for those relevant to psychology, see Byrne, Stewart, & Lee, 2004; to education, see Benson and Bandalos (1992); and to medicine, see Francis, Fletcher, and Rourke (1988).

THE HYPOTHESIZED MODEL

The originally hypothesized model (shown in chap. 6) was tested and modified based on data from the calibration sample (Sample A) of high school teachers. The final best-fitting model for this sample yielded the following goodness-of-fit statistics: S-B$\chi^2_{(422)} = 812.691$; SRMR = .040; *CFI = .960; *RMSEA = .036 with 90% C.I. = .032, .040; it is graphically described in Fig. 8.1. The task now is to determine if this final model replicates over Sample B, the validation group of high school teachers.

THE EQS INPUT FILE

In chapter 7, we tested for the equivalence of a measuring instrument across two different panels of school teachers: elementary and secondary. We first established a baseline model of best fit for each group separately, prior to testing for invariance across groups. However, in cross-validating either a CFA or full structural equation model across groups, the only baseline model determined is for the calibration group. In the present case, this baseline model represents the final best-fitting

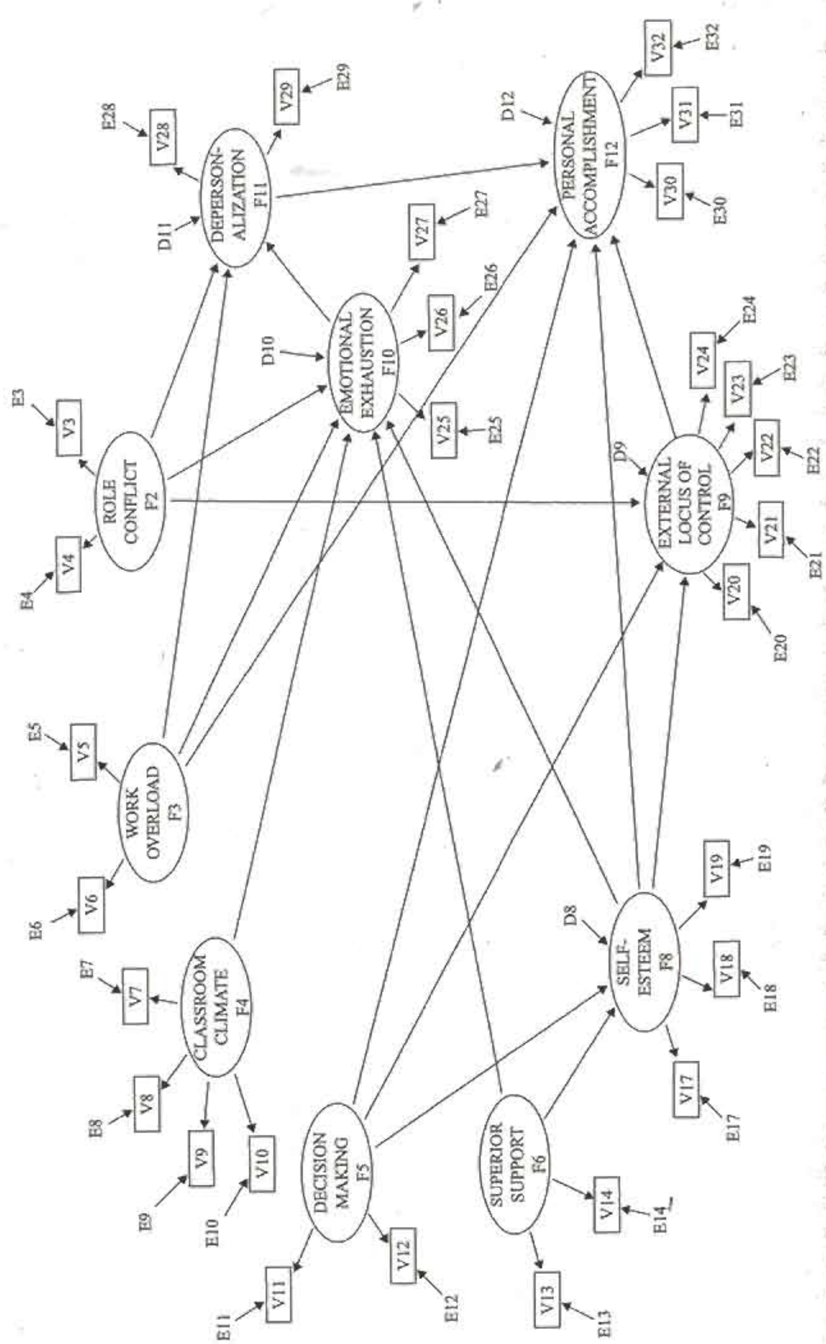

FIG. 8.1. Hypothesized multigroup model of causal structure reflecting determinants of teacher burnout.

model as determined from the analyses presented in chapter 6 and as shown in Fig. 8.1.[1] The rationale here is that the focus in testing across the calibration and validation groups is to determine the extent to which this final model replicates across a second independent sample from the same population.

Consistent with this conceptual rationale, model specification for the validation group should be identical to that for the calibration group. Although space constraints only permit the presentation of EQS input for the calibration group, the exact same specifications are made for the validation group, including two cross-loadings, one error covariance, and all specified start values. This partial file is shown in Table 8.1.

In Table 8.1, we turn first to the /TITLE and /SPECIFICATIONS paragraphs pertinent to Group 1, the calibration group. There is only one major difference in the /SPECIFICATIONS paragraph from the one presented for the same sample of high school teachers in Table 6.1: the statement "Groups = 2." As explained in chapter 7, this statement tells the program that analyses are to be based on two groups. In the same two paragraphs for Group 2, the validation group, observe three (albeit expected) differences in model specification: sample size, data filename, and absence of deleted cases. As noted in Table 8.1, other than these specification differences, the /LABELS, /EQUATIONS, /VARIANCES, and /COVARIANCES paragraphs for the validation group remain identical to those for the calibration group.

In the /CONSTRAINTS paragraph, all factor and cross-loadings, all structural paths, and one error covariance are constrained equal across the two groups. Although the specification of an equality constraint pertinent to the error covariance might be perceived as somewhat excessive (which it is), it is of interest and important to determine if this systematic error holds across the calibration/validation samples. Indeed, the present set of equality constraints represents a very rigorous although necessary test of multigroup invariance. Some researchers might contend that in the case of a full causal model only the structural paths are of interest; however, I maintains that to ensure meaningful and credible interpretation of findings bearing on equality of the structural paths, it is important to know that the measurement parameters (specifically the factor loadings) are operating in the same way for both groups under study.

[1] With reference to the measurement model, the two cross-loadings (i.e., V12→F6 and V30→F10) and one error covariance (i.e., E26, E25) are not included here in the interest of clarity.

TABLE 8.1
EQS Input for Test of Invariant Causal Structure Across Calibration and Validation Samples of Secondary School Teachers

```
/TITLE
 CROSS-VALIDATION OF FINAL BURNOUT MODEL FOR SECONDARY TCHRS "CVBURNHSF"
 GROUP 1: CALIBRATION GROUP
/SPECIFICATIONS
 CASE=716; VAR=32; ME=ML,ROBUST; MA=RAW; FO='(19F4.2/13F4.2)'; Groups=2;
 DATA='C:\EQS61\Files\Books\Data\secind1.ess'; DEL=440,77;
/LABELS
 V1=ROLEA1; V2=ROLEA2; V3=ROLEC1; V4=ROLEC2; V5=WORK1; V6=WORK2; V7=CLIMATE1;
 V8=CLIMATE2; V9=CLIMATE3; V10=CLIMATE4; V11=DEC1; V12=DEC2; V13=SSUP1; V14=SSUP2;
 V15=PSUP1; V16=PSUP2; V17=SELF1; V18=SELF2; V19=SELF3; V20=XLOC1; V21=XLOC2;
 V22=XLOC3; V23=XLOC4; V24=XLOC5; V25=EE1; V26=EE2; V27=EE3; V28=DP1; V29=DP2;
 V30=PA1; V31=PA2; V32=PA3; F1=ROLEA; F2=ROLEC; F3=WORK; F4=CLIMATE; F5=DEC; F6=SSUP;
 F7=PSUP; F8=SELF; F9=XLOC; F10=EE; F11=DP; F12=PA;
/EQUATIONS
  V1  =  1.000 F1  + 1.000 E1 ;
  V2  =  1.308*F1  + 1.000 E2 ;
  V3  =  1.000 F2  + 1.000 E3 ;
  V4  =  1.253*F2  + 1.000 E4 ;
  V5  =  1.000 F3  + 1.000 E5 ;
  V6  =   .708*F3  + 1.000 E6 ;
  V7  =  1.000 F4  + 1.000 E7 ;
  V8  =  1.664*F4  + 1.000 E8 ;
  V9  =   .987*F4  + 1.000 E9 ;
  V10 =  1.385*F4  + 1.000 E10 ;
  V11 =  1.000 F5  + 1.000 E11 ;
  V12 =   .242*F5  + .898*F6 + 1.000 E12 ;
  V13 =  1.000 F6  + 1.000 E13 ;
  V14 =  1.101*F6  + 1.000 E14 ;
  V15 =  1.000 F7  + 1.000 E15 ;
  V16 =  1.060*F7  + 1.000 E16 ;
  V17 =  1.000 F8  + 1.000 E17 ;
  V18 =  1.263*F8  + 1.000 E18 ;
  V19 =  1.417*F8  + 1.000 E19 ;
  V20 =  1.000 F9  + 1.000 E20 ;
  V21 =   .911*F9  + 1.000 E21 ;
  V22 =  1.060*F9  + 1.000 E22 ;
  V23 =   .957*F9  + 1.000 E23 ;
  V24 =  1.246*F9  + 1.000 E24 ;
  V25 =  1.000 F10  + 1.000 E25 ;
  V26 =  1.052*F10  + 1.000 E26 ;
  V27 =  1.219*F10  + 1.000 E27 ;
  V28 =  1.000 F11  + 1.000 E28 ;
  V29 =   .920*F11  + 1.000 E29 ;
  V30 =  -.171*F10  + 1.000 F12 + 1.000 E30 ;
  V31 =  1.234*F12  + 1.000 E31 ;
  V32 =  1.179*F12  + 1.000 E32 ;
  F8  =  .363*F5   -.114*F6   + 1.000 D8 ;
  F9  =  -.262*F5  + *F8 + *F2 + 1.000 D9 ;
  F10 =  -.128*F2 +  .745*F3    -.898*F4  + *F6 + *F8 + 1.000 D10 ;
  F11 =  .549*F10 +  .147*F2  + *F3 + 1.000 D11 ;
  F12 =  .512*F8    -.188*F9  -.195*F11 + *F5 + *F3 + 1.000 D12 ;
/VARIANCES
 F1= .388* ; F2= .676* ; F3= .897* ; F4= .097* ; F5= .566* ; F6= 1.146* ; F7= .622* ;
 E1= .446* ; E2= .321* ; E3= .612* ; E4= .543* ; E5= .561* ; E6= .690* ; E7= .182* ; E8= .138* ;
 E9= .146* ; E10= .323* ; E11= .501* ; E12= .541* ; E13= .452* ; E14= .171* ; E15= .330* ; E16= .154*;
 E17= .072* ; E18= .070* ; E19= .075* ; E20= .211* ; E21= .275* ; E22= .128* ; E23= .235* ; E24= .171* ;
 E25= .688* ; E26= .426* ; E27= .212* ; E28= .246* ; E29= .671* ; E30= .301* ; E31= .308* ; E32= .382* ;
 D8= .077* ; D9= .127* ; D10= .493* ; D11= .628* ; D12= .270* ;
```

(Continued)

TABLE 8.1
(Continued)

```
/COVARIANCES
 F2,F1 = 389* ; F3,F1 = .408* ; F3,F2 = .720* ; F4,F1 = -.038* ; F4,F2 = -.075* ; F4,F3 = -.062* ; F5,F1 = -.382* ;
 F5,F2 = -.493* ; F5,F3 = -.550* ; F5,F4 = .092* ; F6,F1 = -.358* ; F6,F2 = -.497* ; F6,F3 = -.484* ; F6,F4 = .100* ;
 F6,F5 = .636* ; F7,F1 = -.227* ; F7,F2 = -.255* ; F7,F3 = -.286* ; F7,F4 = .049* ; F7,F5 = .417* ; F7,F6 = .391* ;
 E26,E25 = .259* ;
/PRINT
 FIT=ALL;
/END
/TITTLE
 GROUP 2: VALIDATION GROUP
/SPECIFICATIONS
 CASE=714; VAR=32; ME=ML,ROBUST; MA=RAW; FO='(19F4.2/13F4.2)';
 DATA='C:\EQS61\Files\Books\Data\secind2I.ess';
/LABELS
/EQUATIONS        }
/VARIANCES        }  Same as for Calibration group
/COVARIANCES      }
/CONSTRAINTS
 (1,V2,F1) = (2,V2,F1);
 (1,V4,F2) = (2,V4,F2);
 (1,V6,F3) = (2,V6,F3);
 (1,V8,F4) = (2,V8,F4);
 (1,V9,F4) = (2,V9,F4);
 (1,V10,F4) = (2,V10,F4);
 (1,V12,F5) = (2,V12,F5);
 (1,V12,F6) = (2,V12,F6);
 (1,V14,F6) = (2,V14,F6);
 (1,V16,F7) = (2,V16,F7);
 (1,V18,F8) = (2,V18,F8);
 (1,V19,F8) = (2,V19,F8);
 (1,V21,F9) = (2,V21,F9);
 (1,V22,F9) = (2,V22,F9);
 (1,V23,F9) = (2,V23,F9);
 (1,V24,F9) = (2,V24,F9);
 (1,V26,F10) = (2,V26,F10);
 (1,V27,F10) = (2,V27,F10);
 (1,V29,F11) = (2,V29,F11);
 (1,V30,F10) = (2,V30,F10);
 (1,V31,F12) = (2,V31,F12);
 (1,V32,F12) = (2,V32,F12);
 (1,F8,F5) = (2,F8,F5);
 (1,F8,F6) = (2,F8,F6);
 (1,F9,F5) = (2,F9,F5);
 (1,F9,F8) = (2,F9,F8);
 (1,F9,F2) = (2,F9,F2);
 (1,F10,F2) = (2,F10,F2);
 (1,F10,F3) = (2,F10,F3);
 (1,F10,F4) = (2,F10,F4);
 (1,F10,F6) = (2,F10,F6);
 (1,F10,F8) = (2,F10,F8);
 (1,F11,F10) = (2,F11,F10);
 (1,F11,F2) = (2,F11,F2);
 (1,F11,F3) = (2,F11,F3);
 (1,F12,F8) = (2,F12,F8);
 (1,F12,F9) = (2,F12,F9);
 (1,F12,F11) = (2,F12,F11);
 (1,F12,F5) = (2,F12,F5);
 (1,F12,F3) = (2,F12,F3);
 (1,E26,E25) = (2,E26,E25);
/LMTEST
/END
```

THE EQS OUTPUT FILE

Let's turn now to Table 8.2 where the goodness-of-fit results related to the test for multigroup invariance are presented. Somewhat surprisingly (given the rigor of the constraints), these findings reveal a remarkably well-fitting multigroup model (S-B$\chi^2_{(885)}$ = 1623.682; SRMR = .043; *CFI = .961; *RMSEA = .024 with 90% C.I. = .022, .026). Indeed, these results speak well for the general equivalence of the model specifications across calibration and validation samples.

We turn next to Table 8.3 where results from the LM Test of equality constraints are displayed. As explained in chapter 7, in determining evidence of noninvariance based on the LM Test of equality constraints, we look for univariate incremental χ^2 values with probability values < .05. In reviewing these values here, we find two such cases. The one with the smaller probability (Constraint 31; F10, F6) identifies the structural path flowing from Superior Support to Emotional Exhaustion (SSUP → EE); the second (Constraint 4; V8, F4) pinpoints the factor loading of Climate2 onto Factor 4 (Classroom Climate) as being noninvariant across the calibration and validation samples of high school teachers.

In summary, based on these findings we can conclude that with the exception of one structural path flowing from Superior Support to Emotional Exhaustion (SSUP → EE) and the factor loading of V8 (Climate2) onto Factor 4 (Classroom Climate), the hypothesized causal pattern of organizational and personality factor determinants of burnout shown in Fig. 8.1 are equivalent across two independent samples of high school teachers. That the analyses did not yield completely invariant parameters across groups is somewhat disappointing. Nonetheless, given the rigor of the equality constraints imposed, these findings speak well for the cross-validation of this causal structure as it relates to high school teachers.

TABLE 8.2

Selected EQS Output for Test of Invariant Causal Structure: Goodness-of-Fit Statistics

```
GOODNESS OF FIT SUMMARY FOR METHOD = ML

CHI-SQUARE =      1807.331 BASED ON     885 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS        .00000

FIT INDICES
-----------
BENTLER-BONETT     NORMED FIT INDEX =        .924
BENTLER-BONETT NON-NORMED FIT INDEX =        .955
COMPARATIVE FIT INDEX (CFI)         =        .960
ROOT MEAN-SQUARE RESIDUAL (RMR)     =        .037
STANDARDIZED RMR                    =        .043
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)     =        .027
90% CONFIDENCE INTERVAL OF RMSEA  (          .025,          .029)

GOODNESS OF FIT SUMMARY FOR METHOD = ROBUST

SATORRA-BENTLER SCALED CHI-SQUARE =     1623.6828 ON     885 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS        .00000

FIT INDICES
-----------
BENTLER-BONETT     NORMED FIT INDEX =        .919
BENTLER-BONETT NON-NORMED FIT INDEX =        .957
COMPARATIVE FIT INDEX (CFI)         =        .961
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)     =        .024
90% CONFIDENCE INTERVAL OF RMSEA  (          .022,          .026)
```

TABLE 8.3

Selected EQS Output for Test of Invariant Causal Structure: LM Test Results

```
LAGRANGE MULTIPLIER TEST (FOR RELEASING CONSTRAINTS)
CONSTRAINTS TO BE RELEASED ARE:

        CONSTRAINTS FROM GROUP 2

        CONSTR:   1   (1,V2,F1)-(2,V2,F1)=0;
        CONSTR:   2   (1,V4,F2)-(2,V4,F2)=0;
        CONSTR:   3   (1,V6,F3)-(2,V6,F3)=0;
        CONSTR:   4   (1,V8,F4)-(2,V8,F4)=0;
        CONSTR:   5   (1,V9,F4)-(2,V9,F4)=0;
        CONSTR:   6   (1,V10,F4)-(2,V10,F4)=0;
        CONSTR:   7   (1,V12,F5)-(2,V12,F5)=0;
        CONSTR:   8   (1,V12,F6)-(2,V12,F6)=0;
        CONSTR:   9   (1,V14,F6)-(2,V14,F6)=0;
        CONSTR:  10   (1,V16,F7)-(2,V16,F7)=0;
        CONSTR:  11   (1,V18,F8)-(2,V18,F8)=0;
        CONSTR:  12   (1,V19,F8)-(2,V19,F8)=0;
        CONSTR:  13   (1,V21,F9)-(2,V21,F9)=0;
        CONSTR:  14   (1,V22,F9)-(2,V22,F9)=0;
        CONSTR:  15   (1,V23,F9)-(2,V23,F9)=0;
        CONSTR:  16   (1,V24,F9)-(2,V24,F9)=0;
        CONSTR:  17   (1,V26,F10)-(2,V26,F10)=0;
        CONSTR:  18   (1,V27,F10)-(2,V27,F10)=0;
        CONSTR:  19   (1,V29,F11)-(2,V29,F11)=0;
```

(Continued)

TABLE 8.3
(Continued)

```
LAGRANGE MULTIPLIER TEST (FOR RELEASING CONSTRAINTS)
CONSTRAINTS TO BE RELEASED ARE:
        CONSTRAINTS FROM GROUP 2
        CONSTR:  20   (1,V30,F10)-(2,V30,F10)=0;
        CONSTR:  21   (1,V31,F12)-(2,V31,F12)=0;
        CONSTR:  22   (1,V32,F12)-(2,V32,F12)=0;
        CONSTR:  23   (1,F8,F5)-(2,F8,F5)=0;
        CONSTR:  24   (1,F8,F6)-(2,F8,F6)=0;
        CONSTR:  25   (1,F9,F5)-(2,F9,F5)=0;
        CONSTR:  26   (1,F9,F8)-(2,F9,F8)=0;
        CONSTR:  27   (1,F9,F2)-(2,F9,F2)=0;
        CONSTR:  28   (1,F10,F2)-(2,F10,F2)=0;
        CONSTR:  29   (1,F10,F3)-(2,F10,F3)=0;
        CONSTR:  30   (1,F10,F4)-(2,F10,F4)=0;
        CONSTR:  31   (1,F10,F6)-(2,F10,F6)=0;
        CONSTR:  32   (1,F10,F8)-(2,F10,F8)=0;
        CONSTR:  33   (1,F11,F10)-(2,F11,F10)=0;
        CONSTR:  34   (1,F11,F2)-(2,F11,F2)=0;
        CONSTR:  35   (1,F11,F3)-(2,F11,F3)=0;
        CONSTR:  36   (1,F12,F8)-(2,F12,F8)=0;
        CONSTR:  37   (1,F12,F9)-(2,F12,F9)=0;
        CONSTR:  38   (1,F12,F11)-(2,F12,F11)=0;
        CONSTR:  39   (1,F12,F5)-(2,F12,F5)=0;
        CONSTR:  40   (1,F12,F3)-(2,F12,F3)=0;
        CONSTR:  41   (1,E26,E25)-(2,E26,E25)=0;
```

	CUMULATIVE MULTIVARIATE STATISTICS				UNIVARIATE INCREMENT	
STEP	PARAMETER	CHI-SQUARE	D.F.	PROBABILITY	CHI-SQUARE	PROBABILITY
1	CONSTR: 31	6.539	1	.011	6.539	.011
2	CONSTR: 4	11.279	2	.004	4.740	.029
3	CONSTR: 18	14.852	3	.002	3.574	.059
4	CONSTR: 39	18.529	4	.001	3.676	.055
5	CONSTR: 7	21.725	5	.001	3.196	.074
6	CONSTR: 30	24.251	6	.000	2.527	.112
7	CONSTR: 11	26.139	7	.000	1.888	.169
8	CONSTR: 35	28.051	8	.000	1.912	.167
9	CONSTR: 14	29.902	9	.000	1.851	.174
10	CONSTR: 16	33.172	10	.000	3.269	.071
11	CONSTR: 27	34.493	11	.000	1.321	.250
12	CONSTR: 13	36.429	12	.000	1.937	.164

9

Application 7: Testing for Latent Mean Differences Based on a First-Order CFA Model

In the years since the printing of my first EQS book in 1994, there has been a steady albeit moderate increase in reported findings from tests for multigroup equivalence. Review of the SEM literature, however, reveals that most tests for invariance have been based on the analysis of covariance structures COVS, as exemplified in chapters 7 and 8. Despite Sörbom's (1974) introduction of the mean and covariance structures (MACS) strategy in testing for latent mean differences 30 years ago, few studies have been designed to test for latent mean differences across groups based on real (as opposed to simulated) data (see, e.g., Aiken, Stein, & Bentler, 1994; Byrne, 1988b; Cooke et al., 2001; Little, 1997; Marsh & Grayson, 1994; Reise et al., 1993; and Widaman & Reise, 1997). This chapter introduces you to basic concepts associated with the analysis of latent mean structures and walks you through an application that tests for the invariance of latent means across two different cultural groups. Specifically, we test for differences in the latent means of four nonacademic self-concepts (SCs)—Physical SC (Appearance), Physical SC (Ability), Social SC (Peers), and Social SC (Parents)—across Australian and Nigerian adolescents; these constructs comprise the four nonacademic SC components of the Self-Description Questionnaire I (SDQ-I; Marsh, 1992). The present application is taken from a study by Byrne and Watkins (2003) but extends this previous work in two ways: (a) analyses are based on MACS rather than only on COVS, and (b) analyses address the issue of missing data with respect to the Nigerian sample (data were complete for the Australian sample).