TABLE 3.3

EQS Output for Initially Hypothesized Four-Factor Model: Specification and Analysis Summary

COVARIANCE MATRIX TO BE ANALYZED: 16 VARIABLES (SELECTED FROM 46 VARIABLES)
BASED ON 265 CASES.

		SDQ2N01 V 25	SDQ2N13 V 26	SDQ2N25 V 27	SDQ2N37 V 28	SDQ2N04 V 29
SDQ2N01	V 25	1.818				
SDQ2N13	V 26	.684	1.845			
SDQ2N25	V 27	.752	.503	1.505		
SDQ2N37	V 28	.476	.656	.430	1.311	
SDQ2N04	V 29	.420	.665	.301	.470	1.963
SDQ2N16	V 30	.466	.554	.394	.579	.752
SDQ2N28	V 31	.377	.478	.349	.541	.692
SDQ2N40	V 32	.452	.542	.434	.663	.516
SDQ2N10	V 33	.351	.411	.253	.324	.595
SDQ2N22	V 34	.266	.343	.258	.301	.371
SDQ2N34	V 35	.363	.463	.284	.314	.231
SDQ2N46	V 36	.379	.374	.174	.225	.349
SDQ2N07	V 37	.562	.514	.544	.582	.885
SDQ2N19	V 38	.497	.566	.348	.477	.727
SDQ2N31	V 39	.576	.653	.579	.716	.882
SDQ2N43	V 40	.274	.474	.332	.345	.493

		SDQ2N16 V 30	SDQ2N28 V 31	SDQ2N40 V 32	SDQ2N10 V 33	SDQ2N22 V 34
SDQ2N16	V 30	1.539				
SDQ2N28	V 31	.948	1.775			
SDQ2N40	V 32	.878	.929	1.848		
SDQ2N10	V 33	.583	.519	.639	1.327	
SDQ2N22	V 34	.580	.507	.577	.586	1.190
SDQ2N34	V 35	.189	.080	.327	.512	.348
SDQ2N46	V 36	.473	.451	.605	.545	.527
SDQ2N07	V 37	1.022	.978	.890	.246	.246
SDQ2N19	V 38	1.002	1.063	1.039	.350	.366
SDQ2N31	V 39	1.075	1.124	1.017	.341	.371
SDQ2N43	V 40	.746	.747	.837	.226	.187

		SDQ2N34 V 35	SDQ2N46 V 36	SDQ2N07 V 37	SDQ2N19 V 38	SDQ2N31 V 39
SDQ2N34	V 35	2.901				
SDQ2N46	V 36	.454	1.682			
SDQ2N07	V 37	-.328	.213	3.173		
SDQ2N19	V 38	-.080	.161	2.003	2.870	
SDQ2N31	V 39	-.123	.331	2.247	1.763	2.466
SDQ2N43	V 40	-.067	.347	1.435	1.444	1.433

		SDQ2N43 V 40
SDQ2N43	V 40	1.962

BENTLER—WEEKS STRUCTURAL REPRESENTATION:

NUMBER OF DEPENDENT VARIABLES = 16
DEPENDENT V'S : 25 26 27 28 29 30 31 32 33 34
DEPENDENT V'S : 35 36 37 38 39 40

NUMBER OF INDEPENDENT VARIABLES = 20
INDEPENDENT F'S : 1 2 3 4
INDEPENDENT E'S : 25 26 27 28 29 30 31 32 33 34
INDEPENDENT E'S : 35 36 37 38 39 40

NUMBER OF FREE PARAMETERS = 38
NUMBER OF FIXED NONZERO PARAMETERS = 20

PARAMETER ESTIMATES APPEAR IN ORDER,
NO SPECIAL PROBLEMS WERE ENCOUNTERED DURING OPTIMIZATION.

(For an extensive explanation of empirical identification [and underidentification], see Bollen, 1989a; Kenny, 1979; Kline, 1998; Maruyama, 1998; Rindskopf, 1984; and Wothke, 1993.) The second situation results from the presence of boundary parameters—those with values close to the boundary of admissible values; typical examples are correlation estimates greater than 1.00 and variance estimates that are zero or some negative value. In contrast to LISREL, which places no constraints on these parameters, EQS forces them to be held to a boundary value (i.e., 1.00 or zero). As such, the presence of boundary parameters in EQS generates one of two Condition Code messages: (a) "CONSTRAINED AT UPPER BOUND" or (b) "CONSTRAINED AT LOWER BOUND." (For greater elaboration on the cause and alternative approaches to addressing these difficulties, see Bentler, 2005; Bentler & Chou, 1987; Bollen, 1989a; and Rindskopf, 1984.)

Model Assessment. Of primary interest in SEM is the extent to which a hypothesized model "fits" or, in other words, adequately describes the sample data. Given findings of an inadequate goodness-of-fit, the next logical step is to detect the source of misfit in the model. Ideally, evaluation of model fit should derive from a variety of perspectives based on several criteria that can assess model fit. In particular, these criteria focus on the adequacy of (a) the model as a whole, and (b) the individual parameter estimates.

Model as a Whole

Before turning to this section of the EQS output, it is worthwhile to review five important aspects of fitting hypothesized models: (a) the rationale on which the model-fitting process is based, (b) the issue of statistical significance, (c) the estimation process, (d) the residual covariance matrices and (e) the goodness-of-fit statistics.

The Model-Fitting Process. In chap. 1, I presented a general description of this process and noted that the primary task is to determine the goodness-of-fit between the hypothesized model and the sample data. In other words, the researcher specifies a model and then uses the sample data to test the model.

With a view to helping you gain a better understanding of the material to be presented next in the output file, let's take a few moments to recast this model-fitting process within a more formalized framework. As such, let S represent the sample covariance matrix (of observed variable scores), Σ (sigma) the population covariance matrix, and θ (theta) a vector that comprises the model parameters. As such, $\Sigma(\theta)$ represents the restricted covariance matrix implied by the model (i.e., the specified structure of the hypothesized model). In SEM, the null hypothesis (H_0) being tested is that the postulated model holds in the population [i.e., $\Sigma = \Sigma(\theta)$]. In contrast to traditional statistical procedures, however, the researcher hopes *not* to reject H_0 (but see MacCallum, Browne, & Sugarawa, 1996, for proposed changes to this hypothesis-testing strategy).

The issue of statistical significance. The rationale underlying the practice of statistical significance testing has generated a plethora of criticism over at least the past four decades. Indeed, Cohen (1994) noted that, despite Rozeboom's (1960) admonition 45 years ago that "the statistical folkways of a more primitive past continue to dominate the local scene" (p. 417), this dubious practice still persists. (For an array of supportive as well as opposing views with respect to this article by a number of researchers, see the *American Psychologist* [1995], *50*, 1098–1103.) In light of this historical bank of criticism, together with the current pressure by methodologists to cease this traditional ritual (see, e.g., Cohen, 1994; Kirk, 1996; Schmidt, 1996; and Thompson, 1996), the Board of Scientific Affairs for the American Psychological Association appointed a task force to study the feasibility of phasing out the use of null hypothesis testing procedures, as described in course texts and reported in journal articles. Consequently, the end of statistical significance testing relative to traditional statistical methods may soon be a reality.

Statistical significance testing with respect to the analysis of covariance structures, however, is somewhat different in that it is driven by degrees of freedom involving the number of elements in the sample covariance matrix and the number of parameters to be estimated. Nonetheless, it is interesting that many of the issues raised with respect to the traditional statistical methods (e.g., practical significance, importance of confidence intervals, and importance of replication) have long been addressed in SEM applications. Indeed, it was this very issue of practical "nonsignificance" in model testing that led Bentler and Bonett (1980) to develop one of the first subjective indexes of fit (i.e., the NFI). Their work subsequently spawned the development of numerous additional practical indexes of fit, many of which are included in the EQS output shown in Table 3.5. Likewise, the early work of Steiger (1990; and Steiger & Lind, 1980) precipitated the call for use of confidence intervals in the reporting of SEM findings (see, e.g., MacCallum et al., 1996). Finally, the classic paper by Cliff (1983) denouncing the proliferation of post hoc model-fitting and criticizing the apparent lack of concern for the dangers of overfitting models to trivial effects arising from capitalization on chance factors, spirited the development of evaluation indexes (Browne & Cudeck, 1989; and Cudeck & Browne, 1983), as well as a general call for the increased use of cross-validation procedures (see, e.g., MacCallum et al., 1992, 1994).

The estimation process. The primary focus of the estimation process in SEM is to yield parameter values such that the discrepancy (i.e., residual) between the sample covariance matrix S and the population covariance matrix implied by the model $[\Sigma(\theta)]$ is minimal. This objective is achieved by minimizing a discrepancy function, $F[S, \Sigma(\theta)]$, such that its minimal value (F_{min}) reflects the point in the estimation process where the discrepancy between S and $\Sigma(\theta)$ is least $[S - \Sigma(\theta) = \text{minimum}]$. Taken together, then, F_{min} serves as a measure of the extent to which S differs from $\Sigma(\theta)$; any discrepancy between the two is captured by the residual covariance matrix. In EQS, information related to these residuals is presented first, followed by the global goodness-of-fit indexes. Table 3.4 summarizes the residual covariance matrices related to the hypothesized model.

TABLE 3.4

EQS Output for Initially Hypothesized Four-Factor Model: Residuals

RESIDUAL COVARIANCE MATRIX (S-SIGMA):

		SDQ2N01 V 25	SDQ2N13 V 26	SDQ2N25 V 27	SDQ2N37 V 28	SDQ2N04 V 29
SDQ2N01	V 25	.000				
SDQ2N13	V 26	.018	.000			
SDQ2N25	V 27	.229	-.063	.000		
SDQ2N37	V 28	-.098	.033	-.059	.000	
SDQ2N04	V 29	.003	.214	-.053	.080	.000
SDQ2N16	V 30	-.067	-.022	-.058	.081	.031
SDQ2N28	V 31	-.142	-.084	-.093	.056	-.010
SDQ2N40	V 32	-.072	-.026	-.012	.173	-.194
SDQ2N10	V 33	-.005	.025	-.050	-.009	.129
SDQ2N22	V 34	-.051	.000	-.011	.004	-.043
SDQ2N34	V 35	.124	.204	.081	.091	-.080
SDQ2N46	V 36	.079	.049	-.081	-.056	-.044
SDQ2N07	V 37	-.075	-.177	.002	-.013	.009
SDQ2N19	V 38	-.039	-.015	-.108	-.024	-.010
SDQ2N31	V 39	-.031	-.005	.062	.149	.048
SDQ2N43	V 40	-.143	.021	-.024	-.046	-.081

		SDQ2N16 V 30	SDQ2N28 V 31	SDQ2N40 V 32	SDQ2N10 V 33	SDQ2N22 V 34
SDQ2N16	V 30	.000				
SDQ2N28	V 31	.050	.000			
SDQ2N40	V 32	-.028	.044	.000		
SDQ2N10	V 33	-.012	-.061	.053	.000	
SDQ2N22	V 34	.050	-.009	.055	-.010	.000
SDQ2N34	V 35	-.210	-.309	-.066	.062	-.051
SDQ2N46	V 36	-.029	-.039	.111	-.021	.025
SDQ2N07	V 37	-.098	-.114	-.213	-.085	-.049
SDQ2N19	V 38	.060	.144	.111	.071	.118
SDQ2N31	V 39	.008	.084	-.034	.025	.090
SDQ2N43	V 40	.012	.031	.114	.009	-.007

		SDQ2N34 V 35	SDQ2N46 V 36	SDQ2N07 V 37	SDQ2N19 V 38	SDQ2N31 V 39
SDQ2N34	V 35	.000				
SDQ2N46	V 36	.075	.000			
SDQ2N07	V 37	-.551	-.067	.000		
SDQ2N19	V 38	-.267	-.075	.055	.000	
SDQ2N31	V 39	-.335	.065	.042	-.092	.000
SDQ2N43	V 40	-.213	.164	-.082	.168	-.012

		SDQ2N43 V 40
SDQ2N43	V 40	.000

AVERAGE ABSOLUTE COVARIANCE RESIDUALS = .0684
AVERAGE OFF-DIAGONAL ABSOLUTE COVARIANCE RESIDUALS = .0775

(Continued)

TABLE 3.4
(Continued)

STANDARDIZED RESIDUAL MATRIX:

	SDQ2N01 V 25	SDQ2N13 V 26	SDQ2N25 V 27	SDQ2N37 V 28	SDQ2N04 V 29
SDQ2N01 V 25	.000				
SDQ2N13 V 26	.010	.000			
SDQ2N25 V 27	.138	-.038	.000		
SDQ2N37 V 28	-.064	.021	-.042	.000	
SDQ2N04 V 29	.002	.112	-.031	.050	.000
SDQ2N16 V 30	-.040	-.013	-.038	.057	.018
SDQ2N28 V 31	-.079	-.046	-.057	.036	-.006
SDQ2N40 V 32	-.039	-.014	-.007	.111	-.102
SDQ2N10 V 33	-.003	.016	-.035	-.007	.080
SDQ2N22 V 34	-.034	.000	-.008	.004	-.028
SDQ2N34 V 35	.054	.088	.039	.047	-.034
SDQ2N46 V 36	.045	.028	-.051	-.038	-.024
SDQ2N07 V 37	-.031	-.073	.001	-.007	.004
SDQ2N19 V 38	-.017	-.006	-.052	-.012	-.004
SDQ2N31 V 39	-.014	-.002	.032	.083	.022
SDQ2N43 V 40	-.076	.011	-.014	-.029	-.041

	SDQ2N16 V 30	SDQ2N28 V 31	SDQ2N40 V 32	SDQ2N10 V 33	SDQ2N22 V 34
SDQ2N16 V 30	.000				
SDQ2N28 V 31	.030	.000			
SDQ2N40 V 32	-.017	.025	.000		
SDQ2N10 V 33	-.009	-.040	.034	.000	
SDQ2N22 V 34	.037	-.006	.037	-.008	.000
SDQ2N34 V 35	-.099	-.136	-.028	.032	-.028
SDQ2N46 V 36	-.018	-.022	.063	-.014	.017
SDQ2N07 V 37	-.044	-.048	-.088	-.042	-.025
SDQ2N19 V 38	.028	.064	.048	.036	.064
SDQ2N31 V 39	.004	.040	-.016	.014	.053
SDQ2N43 V 40	.007	.017	.060	.005	-.004

	SDQ2N34 V 35	SDQ2N46 V 36	SDQ2N07 V 37	SDQ2N19 V 38	SDQ2N31 V 39
SDQ2N34 V 35	.000				
SDQ2N46 V 36	.034	.000			
SDQ2N07 V 37	-.181	-.029	.000		
SDQ2N19 V 38	-.093	-.034	.018	.000	
SDQ2N31 V 39	-.125	.032	.015	-.034	.000
SDQ2N43 V 40	-.089	.090	-.033	.071	-.005

	SDQ2N43 V 40
SDQ2N43 V 40	.000

AVERAGE ABSOLUTE STANDARDIZED RESIDUALS = .0342
AVERAGE OFF-DIAGONAL ABSOLUTE STANDARDIZED RESIDUALS = .0388

TABLE 3.4
(Continued)

LARGEST STANDARDIZED RESIDUALS:

NO.	PARAMETER	ESTIMATE	NO.	PARAMETER	ESTIMATE
1	V37, V35	-.181	11	V40, V35	-.089
2	V27, V25	.138	12	V35, V26	.088
3	V35, V31	-.136	13	V37, V32	-.088
4	V39, V35	-.125	14	V39, V28	.083
5	V29, V26	.112	15	V33, V29	.080
6	V32, V28	.111	16	V31, V25	-.079
7	V32, V29	-.102	17	V40, V25	-.076
8	V35, V30	-.099	18	V37, V26	-.073
9	V38, V35	-.093	19	V40, V38	.071
10	V40, V36	.090	20	V38, V34	.064

DISTRIBUTION OF STANDARDIZED RESIDUALS

```
 ----------------------------------
 !                                 !
80-                                -
 !                                 !
 !              *                  !
 !              *                  !
 !              *  *               !
60-             *  *               -
 !              *  *               !
 !              *  *               !
 !              *  *               !
40-             *  *               -
 !              *  *               !
 !              *  *               !
 !              *  *               !
20-             *  *               -
 !              *  *               !
 !              *  *               !
 !              *  *               !
 !           *  *  *  *            !
 ----------------------------------
    1  2  3  4  5  6  7  8  9  A  B  C
```

	RANGE			FREQ	PERCENT
1	-0.5	-	--	0	.00%
2	-0.4	-	-0.5	0	.00%
3	-0.3	-	-0.4	0	.00%
4	-0.2	-	-0.3	0	.00%
5	-0.1	-	-0.2	4	2.94%
6	0.0	-	-0.1	68	50.00%
7	0.1	-	0.0	61	44.85%
8	0.2	-	0.1	3	2.21%
9	0.3	-	0.2	0	.00%
A	0.4	-	0.3	0	.00%
B	0.5	-	0.4	0	.00%
C	++	-	0.5	0	.00%
	TOTAL			136	100.00%

EACH "*" REPRESENTS 4 RESIDUALS

Residual covariance matrices. Each element in the residual matrix represents the discrepancy between the covariances in $\Sigma(\theta)$ and those in S [i.e., $\Sigma(\theta) - S$]; that is, there is one residual for each pair of observed variables (Jöreskog, 1993). In the case of this hypothesized model, for example, the residual matrix would contain $[(16 \times 17)/2] = 136$ elements. It is worth noting that, as in conventional regression analysis, the residuals are not independent of one another. Thus, any attempts to test them (in the strict statistical sense) would be inappropriate. In essence, only their magnitude is of interest in alerting the researcher to possible areas of model misfit.

EQS provides both unstandardized and standardized residual covariance matrices. Given that a model describes the data well, these residual values should be small and evenly distributed. Large residuals associated with particular parameters indicate their misspecification in the model, thereby affecting the overall model misfit. For both the unstandardized and standardized residual matrices, EQS computes two averages: one based on all elements of the lower triangular matrix, the other ignoring the diagonal elements. Typically, the off-diagonal elements play a more major role in the effect of goodness-of-fit χ^2 statistics (Bentler, 2005). Based on an ordering from large to small, the program then lists the 20 largest standardized residuals and designates which pairs of variables are involved. Finally, a frequency distribution of the standardized residuals is presented. Ideally, this distribution should be symmetric and centered around zero.

Because the fitted residuals are dependent on the unit of measurement of the observed variables, they can be difficult to interpret; thus, their standardized values are typically examined. Standardized residuals are fitted residuals divided by their asymptotically (large sample) standard errors (Jöreskog & Sörbom, 1988). As such, they are analogous to Z-scores and are therefore the easier of the two sets of residual values to interpret. In essence, they represent estimates of the number of standard deviations the observed residuals are from the zero residuals that would exist if model fit were perfect [i.e., $\Sigma(\theta) - S = .0$]. Values >2.58 are considered large (Jöreskog & Sörbom, 1988). In examining the standardized residual information in Table 3.4, we see that the average off-diagonal value is .0342, whereas the largest off-diagonal value is .0388, both of which reflect a very good fit to the data. Finally, a review of the frequency distribution reveals most residual values (94.85%) to fall between −.1 and .1. Of the remaining residuals, 2.94% fall between −0.1 and −0.2 (see values −1.81, .136, .125, .102) and 2.21% fall between 0.1 and 0.2 (see values .138, .112, .111). From this information, we can conclude that although there may be some minimal discrepancy in fit between the hypothesized model and the sample data, overall, the model as a whole appears to be quite well fitting.

The goodness-of-fit statistics. Let's turn now to the goodness-of-fit statistics presented in Table 3.5. Here we find statistics reported for several goodness-of-fit values, all of which relate to the model as a whole.[7] The first of these values represents the INDEPENDENCE CHI-SQUARE statistic ($\chi^2_{(120)} = 1696.728$), as it relates to the Independence model. This model (also termed the *null* model) is so named because it represents complete independence from all variables in the model (i.e., all variables in the model are mutually uncorrelated). Although other baseline models have been proposed (see, e.g., Sobel & Bohrnstedt, 1985), Rigdon (1996) noted that beginning with the work of Tucker and Lewis (1973), the independence baseline model is the one most widely used. Indeed, Bentler and Bonett (1980) argued that in large samples, the independence model serves as a good baseline against which to compare alternative models to evaluate the gain in improved fit. That is, given a sound hypothesized model, one would expect the χ^2 value for the independence model to be extremely high, thereby indicating excessive misfit; such is the case with the present example.

[7]Had the /PRINT paragraph not been included in the input file, only the default goodness-of-fit indexes would have been reported; these include all statistics down to and including the CFI.

TABLE 3.5

EQS Output for Initially Hypothesized Four-Factor Model: Goodness-of-Fit Statistics

```
GOODNESS OF FIT SUMMARY FOR METHOD = ML

INDEPENDENCE MODEL CHI-SQUARE    =    1696.728 ON    120 DEGREES OF FREEDOM

INDEPENDENCE AIC =  1456.72826    INDEPENDENCE CAIC =   907.16068
       MODEL AIC =   -37.48818           MODEL CAIC =  -486.30170

CHI-SQUARE =       158.512 BASED ON       98 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS        .00011

THE NORMAL THEORY RLS CHI-SQUARE FOR THIS ML SOLUTION IS        152.727.

FIT INDICES
-----------
BENTLER-BONETT      NORMED FIT INDEX =       .907
BENTLER-BONETT NON-NORMED FIT INDEX =       .953
COMPARATIVE FIT INDEX (CFI)         =       .962
BOLLEN    (IFI) FIT INDEX           =       .962
MCDONALD  (MFI) FIT INDEX           =       .892
LISREL    GFI   FIT INDEX           =       .933
LISREL    AGFI  FIT INDEX           =       .906
ROOT MEAN-SQUARE RESIDUAL (RMR)     =       .104
STANDARDIZED RMR                    =       .048
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)      =      .048
90% CONFIDENCE INTERVAL OF RMSEA  (          .034,          .062)

                      ITERATIVE SUMMARY

                   PARAMETER
ITERATION          ABS CHANGE          ALPHA              FUNCTION
    1               .584046            .50000             3.03565
    2               .313485           1.00000             1.36061
    3               .076048           1.00000              .62068
    4               .038441           1.00000              .60229
    5               .004573           1.00000              .60056
    6               .003231           1.00000              .60044
    7               .000542           1.00000              .60042
```

Now skip down three lines to the chi-square value reported for the hypothesized four-factor model. The value of 158.512 represents the discrepancy between the unrestricted sample covariance matrix S and the restricted covariance matrix $\Sigma(\theta)$ and, in essence, represents the Likelihood Ratio Test statistic, most commonly expressed as a chi-square (χ^2) statistic. This statistic is equal to $(N - 1)F_{min}$ (i.e., sample size minus 1 multiplied by the minimum fit function) and in large samples is distributed as a central χ^2 with degrees of freedom equal to $1/2(p)(p+1) - t$, where p is the number of observed variables and t is the number of parameters to be estimated (Bollen, 1989a). In general, H_0: $\Sigma = \Sigma(\theta)$ is equivalent to the hypothesis that $\Sigma - \Sigma(\theta) = 0$; the χ^2 test, then, simultaneously tests the extent to which all residuals in $\Sigma - \Sigma(\theta)$ are zero (Bollen, 1989a). Framed somewhat differently, the null hypothesis (H_0) postulates that specification of the factor loadings, factor variances–covariances, and error variances for the model under study are valid; the χ^2 test simultaneously tests the extent to which this specification is true. The probability value associated with χ^2 represents the likelihood of obtaining a χ^2 value that exceeds the χ^2 value when H_0 is true. Thus, the higher the probability associated with χ^2, the closer the fit between the hypothesized model (under H_0) and the perfect fit (Bollen, 1989a).

The test of H_0—that SC is a four-factor structure as depicted in Fig. 3.1—yielded a χ^2 value of 158.512, with 98 degrees of freedom and a probability of less than .0001 ($p < .0001$), thereby suggesting that the fit of the data to the hypothesized model is not entirely adequate. Interpreted literally, this test statistic indicates that given the present data, the hypothesis bearing on SC relations, as summarized in the model, represents an unlikely event (i.e., occurring less than one time in a thousand under the null hypothesis) and should be rejected. The sensitivity of the χ^2 likelihood ratio test to sample size, however, is well known and is addressed shortly.

In addition to furnishing χ^2 statistics for the independence and hypothesized models, EQS provides for the evaluation of both models based on Akaike's (1987) Information Criterion (AIC) and Bozdogan's (1987) consistent version of the AIC (CAIC). Both criteria address the issue of parsimony in the assessment of model fit; that is, statistical goodness-of-fit as well as the number of estimated parameters are taken into account. However, Bozdogan (1987) noted that the AIC carried a penalty only as it related to degrees of freedom (thereby reflecting the number of estimated parameters in the model) and not to sample size; he subsequently proposed the CAIC, which takes sample size into account (Bandalos, 1993). Although both criteria were developed for maximum likelihood (ML) estimation, they are applied to all estimation methods in EQS.

The AIC and CAIC are used in the comparison of two or more models with smaller values representing a better fit of the hypothesized model (Hu & Bentler, 1995). The AIC and CAIC indexes also share the same conceptual framework; as such, they reflect the extent to which parameter estimates from the original sample will cross-validate in future samples (Bandalos, 1993). Returning to the output, we see that the AIC statistic for both the independence and hypothesized models is substantially smaller than the χ^2 statistic.[8]

Before reviewing the remaining goodness-of-fit statistics, let's first return to the issue of χ^2 sensitivity. In particular, both the sensitivity of the Likelihood Ratio Test to sample size and its basis on the central χ^2 distribution, which assumes that the model fits perfectly in the population (i.e., that H_0 is correct), have led to problems of fit that are now widely known. Because the χ^2 statistic equals $(N-1)F_{min}$, this value tends to be substantial when the model does not hold and sample size is large (Jöreskog & Sörbom, 1993). Yet, the analysis of covariance structures is grounded in large sample theory. As such, large samples are critical to obtaining precise parameter estimates, as well as to the tenability of asymptotic

[8]Readers are referred to the EQS manual (Bentler, 2005) for an explanation of the CAIC.

distributional approximations (MacCallum et al., 1996). Thus, findings of well-fitting hypothesized models, where the χ^2 value approximates the degrees of freedom, have proven to be unrealistic in most SEM empirical research. More common are findings of a large χ^2 relative to degrees of freedom, thereby indicating a need to modify the model to better fit the data (Jöreskog & Sörbom, 1993). Thus, results related to the test of the hypothesized model are not unexpected. Indeed, given this problematic aspect of the Likelihood Ratio Test and the fact that postulated models (no matter how good) can only ever fit real-world data approximately and never exactly, MacCallum et al. (1996) proposed changes to the traditional hypothesis-testing approach in covariance structure modeling. (For an extended discussion of these changes, see MacCallum et al., 1996.)

Researchers addressed the χ^2 limitations by developing goodness-of-fit indexes that take a more pragmatic approach to the evaluation process. To this end, the past two decades have witnessed a plethora of newly developed fit indexes as well as unique approaches to the model-fitting process (for reviews, see, e.g., Gerbing & Anderson, 1993; Hu & Bentler, 1995; Marsh, Balla, & McDonald, 1988; and Tanaka, 1993). These criteria, referred to as "subjective," "practical," or "ad hoc" indexes of fit, are now commonly used as adjuncts to the χ^2 statistic. At the time of writing this book, EQS users are able to select from 10 of these indexes, as reproduced in Table 3.5 (but only if the /PRINT ↳ Fit = All is specified in the input file).

The first four fit indexes listed in the output (see Table 3.5) fall into the category of comparative (Browne, MacCallum, Kim, Andersen, & Glaser, 2002) or incremental (Hu & Bentler, 1995, 1999) fit indexes. These indexes measure the proportionate improvement in fit by comparing a hypothesized model with a more restricted, nested baseline model.[9] As discussed previously, the independence (or *null*) model is typically the most commonly used baseline model (Hu & Bentler, 1999; and Rigdon, 1996).

For more than two decades, Bentler and Bonett's (1980) Normed Fit Index (NFI) has been the practical criterion of choice, as evidenced in large part by the current "classic" status of its original paper (Bentler, 1992; and Bentler & Bonett, 1987). However, addressing evidence that the NFI has shown a tendency to underestimate fit in small samples, Bentler (1990) revised the NFI to consider sample size and proposed the Comparative Fit Index (CFI). Values for both the NFI and CFI range from zero to 1.00 and are derived from comparison between the hypothesized and independence models, as described previously. As such, each provides a measure of complete covariation in the data. Although a value >.90 was originally considered representative of a well-fitting model (see Bentler, 1992), a revised cutoff value close to 0.95 has been advised (Hu & Bentler, 1999). Although both indexes of fit

[9]Nested models are hierarchically related to one another in the sense that their parameter sets are subsets of one another (i.e., particular parameters are freely estimated in one model but fixed to zero in a second model) (Bentler & Chou, 1987; and Bollen, 1989a).

are reported in the EQS output, Bentler (1990) suggested that the CFI should be the index of choice. The program also reports the Non-Normed Fit Index (NNFI), a variant of the NFI that takes model complexity into account. Values for the NNFI can exceed those reported for the NFI and can also fall outside the zero to 1.00 range.

As shown in Table 3.5, all three indexes (NFI = .907, NNFI = .953, CFI = .962) were consistent in suggesting that the hypothesized model represented an adequate fit to the data, albeit the value reported for the NFI reflected only a marginally well-fitting model. However, considering the CFI to be the most appropriate index of the three, we consider the fit of this model to be satisfactory.

The Incremental Fit Index (IFI; Bollen, 1989b) represents a derivative of the NFI; as with both the NFI and CFI, the IFI coefficient values range from zero to 1.00, with values close to 0.95 indicating superior fit (see Hu & Bentler, 1999). More specifically, the IFI was developed to address the issues of parsimony and sample size, which were known to be associated with the NFI. As such, its computation is basically the same as the NFI, except that degrees of freedom are considered. Thus, it is not surprising that the finding of IFI = .962 is consistent with that of the CFI in reflecting a well-fitting model.

The next three fit indexes reported in the output file in Table 3.5 (i.e., MFI, GFI, and AGFI) belong to the category of "absolute" fit indexes. In contrast to the previous incremental fit indexes, the absolute fit indexes do not rely on comparison with a reference model to determine the amount of improvement in model fit; rather, they depend only on how well the hypothesized model fits the sample data (Browne et al., 2002; and Hu & Bentler, 1999). Nonetheless, Hu and Bentler (1999, p. 2) noted that "an implicit or explicit comparison may be made to a saturated model that exactly reproduces the sample covariance matrix."[10] The McDonald Fit Index (MFI; McDonald, 1989) represents a normed measure of the centrality parameter that transforms the rescaled noncentrality parameter,[11] which assesses model misfit (Hu & Bentler, 1995). Although the MFI is similar to the RMSEA (described shortly), it does not provide for a fit-per-degree-of-freedom interpretation. The Goodness-of-Fit Index (GFI; Jöreskog & Sörbom, 1984) is

[10]A saturated model is one in which the number of estimated parameters equals the number of data points (i.e., variances and covariances of the observed variables as in the case of the just-identified model). Conceptualized within the framework of a continuum, the saturated (i.e., least restricted) model would represent one extreme endpoint, whereas the independence (the most restricted) model would represent the other; a hypothesized model always represents a point somewhere between the two.

[11]The noncentrality parameter is a fixed parameter with associated degrees of freedom and can be denoted as $\chi^2_{(df,\lambda)}$. Essentially, it functions as a measure of the discrepancy between Σ and $\Sigma(\theta)$ and thus can be regarded as a "population badness-of-fit" (Steiger, 1990). As such, the greater the discrepancy between Σ and $\Sigma(\theta)$, the larger the λ value. It is now easy to see that the central χ^2 statistic is a special case of the noncentral χ^2 distribution when λ=0.0. (For an excellent discussion and graphic portrayal of differences between the central and noncentral χ^2 statistics, see MacCallum et al., 1996.)

a measure of the relative amount of variance and covariance in S that is jointly explained by Σ. The AGFI differs from the GFI only in the fact that it adjusts for the number of degrees of freedom in the specified model. As such, it addresses the issue of parsimony by incorporating a penalty for the inclusion of additional parameters. Although the GFI and AGFI are commonly reported in the SEM literature, Hu and Bentler (1998) recommended against their use as indexes of fit. In addition to being insufficiently and inconsistently sensitive to model misspecification, Marsh et al. (1988) have shown both indexes to be strongly influenced by sample size.

Although values reported for these three absolute fit indexes range from zero to 1.0, the MFI can exceed 1.0 due to sampling error (Browne et al., 2002; and Hu & Bentler, 1995, 1999), and it is possible for both the GFI and AGFI to be negative (Jöreskog and Sörbom, 1993). The latter, of course, should not occur because it would reflect the fact that the model fits worse than no model at all. Although values greater than 0.90 for the GFI and AGFI are considered to represent a well-fitting model, Hu and Bentler (1999) suggested a cutoff score of .89 for the MFI. Based on the MFI, GFI, and AGFI values reported in Table 3.5 (.892, .933, and .906, respectively), albeit cognitive of their deficiencies noted earlier, we can again conclude that the hypothesized model fits the sample data fairly well.

Finally, although the last three indexes listed in Table 3.5 (i.e., RMR, SRMR, and RMSEA) are also categorized as absolute fit indexes, Browne et al. (2002, p. 405) termed them more specifically as "absolute *misfit* indices." Although both sets of indexes depend only on the fit of the hypothesized model, the absolute fit indexes (MFI, GFI, and AGFI in this instance) increase as goodness-of-fit improves, whereas the absolute misfit indexes decrease as goodness-of-fit improves and attain their lower-bound value of zero when the model fits perfectly (Browne et al., 2002).

The root mean square residual (RMR) represents the average residual value derived from the fitting of the variance–covariance matrix for the hypothesized model $\Sigma(\theta)$ to the variance–covariance matrix of the sample data (S). However, because these residuals are relative to the sizes of the observed variances and covariances, they are difficult to interpret. Thus, they are best interpreted in the metric of the correlation matrix (Hu & Bentler, 1995; and Jöreskog & Sörbom, 1989). Therefore, the standardized RMR (SRMR) represents the average value across all standardized residuals and ranges from zero to 1.00; in a well-fitting model, this value is small—say .05 or less. Review of the output in Table 3.5 shows that the unstandardized residual value reported for the hypothesized model is .104, whereas the SRMR value is .048. Given that the SRMR represents the average discrepancy between the observed sample and hypothesized correlation matrices, we can interpret this value as meaning that the model explains the correlations to within an average error of .043 (Hu & Bentler, 1995).

The Root Mean Square Error of Approximation (RMSEA) and the conceptual framework within which it is embedded were first proposed by Steiger and Lind in 1980, yet the RMSEA has only recently been recognized as one of the most informative criteria in covariance structure modeling. The RMSEA considers the error of approximation in the population and asks the question, "How well would the model, with unknown but optimally chosen parameter values, fit the population covariance matrix if it were available?" (Browne & Cudeck, 1993, pp. 137–8). This discrepancy, as measured by the RMSEA, is expressed per degree of freedom, thus making it sensitive to the number of estimated parameters in the model (i.e., the complexity of the model). Values less than .05 indicate good fit, and values as high as .08 represent reasonable errors of approximation in the population (Browne & Cudeck, 1993). MacCallum et al. (1996) elaborated on those cutpoints and noted that RMSEA values ranging from .08 to .10 indicate mediocre fit and those greater than .10 indicate poor fit. Although Hu and Bentler (1999) suggested a value of .06 to indicate good fit between the hypothesized model and the observed data, they cautioned that when sample size is small, the RMSEA tends to over-reject true population models. Noting that these criteria are based solely on subjective judgment and therefore cannot be regarded as infallible or correct, Browne and Cudeck (1993) and MacCallum et al. (1996) nonetheless argued, that they would appear to be more realistic than a requirement of exact fit, where RMSEA = 0.0. (For a generalization of the RMSEA to multiple independent samples, see Steiger, 1998.)

Overall, MacCallum and Austin (2000) strongly recommended routine use of the RMSEA for at least three reasons: (a) it would appear to be adequately sensitive to model misspecification (Hu & Bentler, 1998); (b) commonly used interpretative guidelines would appear to yield appropriate conclusions regarding model quality (Hu & Bentler, 1998, 1999); and (c) it is possible to build confidence intervals around RMSEA values.

Addressing Steiger's (1990) call for the use of confidence intervals to assess the precision of RMSEA estimates, EQS reports a 90% interval around the RMSEA value. In contrast to point estimates of model fit (which do not reflect the imprecision of the estimate), confidence intervals can yield this information, thereby providing the researcher with more assistance in the evaluation of model fit. Thus, MacCallum et al. (1996) strongly urged the use of confidence intervals in practice. Presented with a small RMSEA, albeit a wide confidence interval, a researcher would conclude that the estimated discrepancy value is quite imprecise, thereby negating any possibility of accurately determining the degree of fit in the population. In contrast, a narrow confidence interval would argue for good precision of the RMSEA value in reflecting model fit in the population (MacCallum et al., 1996).

Table 3.5 shows that the RMSEA value for the hypothesized model is .048, with the 90% confidence interval ranging from .034 to .062. Interpretation of the confidence interval indicates that we can be 90% confident that the true RMSEA value in the population will fall within the bounds of .034 and .062, which represents a good degree of precision. Given that (a) the RMSEA point estimate is <.05 (.048), and (b) the upper bound of the 90% interval is .062, which is less

than the value suggested by Browne and Cudeck (1993)—albeit equal to the cut-off value proposed by Hu and Bentler (1999)—we can conclude that the initially hypothesized model fits the data well.

Before leaving this discussion of the RMSEA, it is important to note that confidence intervals can be influenced seriously by sample size as well as model complexity (MacCallum et al., 1996). For example, if sample size is small and the number of estimated parameters is large, the confidence interval will be wide. Given a complex model (i.e., a large number of estimated parameters), a large sample size would be required to obtain a reasonably narrow confidence interval. Conversely if the number of parameters is small, then the probability of obtaining a narrow confidence interval is high, even for samples of rather moderate size (MacCallum et al., 1996).

Having worked your way through these goodness-of-fit measures, you no doubt are feeling totally overwhelmed and wondering what to do with all this information! Although the entire set of fit indexes does not need to be reported, such an array can provide a good sense of how well a model fits the sample data. But how does one choose which indexes are appropriate in the assessment of model fit? Unfortunately, this choice is not a simple one, primarily because particular indexes have been shown to operate somewhat differently given the sample size, estimation procedure, model complexity, and/or violation of the underlying assumptions of multivariate normality and variable independence. Thus, Hu and Bentler (1995) cautioned, that in choosing which goodness-of-fit indexes to use in the assessment of model fit, careful consideration of these critical factors is essential. In reporting results for the remaining applications in this book, goodness-of-fit indexes are limited to the CFI, SRMR, and RMSEA, along with the related χ^2 value and RMSEA 90% confidence interval. Readers who want further elaboration on these goodness-of-fit statistics with respect to their formulae and functions—and/or the extent to which they are affected by sample size, estimation procedures, misspecification, and/or violations of assumptions—are referred to Bandalos, 1993; Bentler & Yuan, 1999; Bollen, 1989a; Browne & Cudeck, 1993; Curran, West, & Finch, 1996; Fan, Thompson, & Wang, 1999; Finch, West, & MacKinnon, 1997; Gerbing & Anderson, 1993; Hu & Bentler, 1995, 1998, 1999; Hu, Bentler, & Kano, 1992; Jöreskog & Sörbom, 1993; La Du & Tanaka, 1989; Marsh et al., 1988; Mulaik, James, van Altine, Bennett, Lind, & Stilwell, 1989; Raykov & Widaman, 1995; Sugawara & MacCallum, 1993; Tomarken & Waller, 2005; Weng & Cheng, 1997; West, Finch, & Curran, 1995; Wheaton, 1987; and Williams & Holahan, 1994. For an annotated bibliography, see Austin & Calderón, 1996.

To finalize this subsection on model assessment, I wish to leave you with this important reminder: global fit indexes alone cannot possibly envelop all that needs to be known about a model to judge the adequacy of its fit to the sample data. As Sobel and Bohrnstedt (1985, p. 158) so cogently stated two decades ago, "Scientific progress could be impeded if fit coefficients (even appropriate ones) are used as the primary criterion for judging the adequacy of a model." They further posited that despite the problematic nature of the χ^2 statistic, exclusive reliance on goodness-of-fit indexes is unacceptable. Indeed, fit indexes provide no guarantee that a model is useful. In fact, it is entirely possible for a model to fit well and still be incorrectly specified (Wheaton, 1987). (For an excellent review of ways by which such a seemingly dichotomous event can happen, see Bentler and Chou, 1987.) Fit indexes yield information bearing only on the model's *lack of fit*. More important, they can in no way reflect the extent to which the model is plausible; *this judgment rests squarely on the shoulders of the researcher*. Thus, assessment of model adequacy must be based on multiple criteria that take into account theoretical, statistical, and practical considerations.

The last piece of information related to overall model fit appearing on the output in Table 3.5 is the ITERATIVE SUMMARY. Here we see a synopsis of the number of iterations required for a convergent solution and the mean absolute change in parameter estimates (PARAMETER ABS CHANGE) associated with each iteration. The best scenario is a situation in which only a few iterations are needed to reach convergence; after the first two or three iterations, the change in parameter estimates stabilizes and remains minimal. As indicated in Table 3.5, this is the case with our CFA model, in which only seven iterations were needed to reach convergence. After the first three iterations, the parameter values remained relatively stable.

At the very worst, the number of iterations exceeds the default value of 30, resulting in nonconvergence; as such, the iterative process terminates and a message warning the user not to trust the output is issued. If this problem is presented, it is unlikely that a simple resubmission of the job with a requested extension in the number of iterations (e.g., /TECHNICAL↳ Iter = 500;) will resolve the dilemma. Rather, the user should look for other means of resolution. In my experience, I have found this situation to be easily solved just by attending to the start values. If start values were not included in the input file, then add them; if they were included, make a few modifications. Given that start values were included, lack of convergence occurs most often due to a wide discrepancy between the start values and actual estimated values related to only a few parameters. A typical example is one in which the start value is positive but the actual estimate is negative. A quick way to determine more appropriate start values is to review the estimates provided with the failed output; despite the fact that many of these estimates may be inaccurate, they can often guide the user to a start value that better approximates the actual estimated value. The most efficient approach to achieving more appropriate start values is to use the RETEST option provided by EQS. This option is introduced and discussed in Chap. 6.

A final point about Table 3.5 is the FUNCTION column. In general, EQS minimizes a fit function and when iterations stop, this value should be at the minimum value, with $\chi^2 = (N - 1) \times$ Function. Within the context of the current model, this formulation gives a χ^2 value of $264 \times 0.60042 = 158.51$.

Assessment of Individual Parameter Estimates

This discussion of model fit assessment has thus far concentrated on the model as a whole. Now, we turn our attention to the fit of individual parameters in the model. There are two aspects of concern: (a) the appropriateness of the estimates, and (b) their statistical significance. Parameter estimates and related information are presented in Table 3.6.

Feasibility of parameter estimates. The initial step in assessing the fit of individual parameters in a model is to determine the viability of their estimated values. Specifically, parameter estimates should exhibit the correct sign and size and be consistent with the underlying theory. Any estimates falling outside the admissable range signal a clear indication that either the model is wrong or the input matrix lacks sufficient information. Examples of parameters exhibiting unreasonable estimates are correlations >1.00, negative variances, and covariance or correlation matrices that are not positive definite.

Appropriateness of standard errors. Another indicator of poor model fit is the presence of standard errors that are excessively large or small. For example, if a standard error approaches zero, the test statistic for its related parameter cannot be defined (Bentler, 2005). Likewise, standard errors that are extremely large indicate parameters that cannot be determined (Jöreskog & Sörbom, 1989). Because standard errors are influenced by the units of measurement in observed and/or latent variables, as well as the magnitude of the parameter estimate itself, no definitive criterion of "small" and "large" has been established (Jöreskog & Sörbom, 1989).

Statistical significance of parameter estimates. The test statistic here represents the parameter estimate divided by its standard error; as such, it operates as a Z-statistic in testing that the estimate is statistically different from zero. Based on an α level of .05, the test statistic needs to be $> \pm 1.96$ before the hypothesis (i.e., that the estimate $= 0.0$) can be rejected. Nonsignificant parameters, with the exception of error variances, can be considered unimportant to the model; in the interest of scientific parsimony, albeit given an adequate sample size, they should be deleted from the model. Conversely nonsignificant parameters can be indicative of a sample size that is too small (Jöreskog, pers. comm., January 1997). Finally, conclusions based on a series of univariate tests, as in this case, may differ from those based on a multivariate test in which a set of parameters is considered simultaneously. Although this multivariate option is available to EQS users via the Wald Test (WTest; Wald, 1943), it is not considered herein due to space constraints but is illustrated in subsequent chapters.

Scanning the output in the printout presented in Table 3.6, we see that the unstandardized estimates are presented first, followed by the standardized solution. Both sets of estimates are presented separately for the measurement equations, the variances, and the covariances. Looking more closely at the unstandardized estimates, we see that for variables SDQ2N01 (V25), SDQ2N04 (V29), SDQ2N10 (V33),

TABLE 3.6
EQS Output for Initially Hypothesized Four-Factor Model: Parameter Estimates

```
MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
 STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.
SDQ2N01 =V25 =   1.000 F1    + 1.000 E25
SDQ2N13 =V26 =   1.084*F1    + 1.000 E26
           .154
          7.027@
SDQ2N25 =V27 =    .851*F1    + 1.000 E27
           .132
          6.437@
SDQ2N37 =V28 =    .935*F1    + 1.000 E28
           .131
          7.117@
SDQ2N04 =V29 =   1.000 F2    + 1.000 E29
SDQ2N16 =V30 =   1.278*F2    + 1.000 E30
           .150
          8.507@
SDQ2N28 =V31 =   1.247*F2    + 1.000 E31
           .154
          8.083@
SDQ2N40 =V32 =   1.259*F2    + 1.000 E32
           .157
          8.037@
SDQ2N10 =V33 =   1.000 F3    + 1.000 E33
SDQ2N22 =V34 =    .889*F3    + 1.000 E34
           .103
          8.643@
SDQ2N34 =V35 =    .670*F3    + 1.000 E35
           .148
          4.528@
SDQ2N46 =V36 =    .843*F3    + 1.000 E36
           .117
          7.212@
SDQ2N07 =V37 =   1.000 F4    + 1.000 E37
SDQ2N19 =V38 =    .841*F4    + 1.000 E38
           .058
         14.471@
SDQ2N31 =V39 =    .952*F4    + 1.000 E39
           .049
         19.475@
SDQ2N43 =V40 =    .655*F4    + 1.000 E40
           .049
         13.273@
```


TABLE 3.6
(Continued)

```
VARIANCES OF INDEPENDENT VARIABLES
-----------------------------------
      V                  F
     ---                ---
              I F1 - F1         .615*
              I                 .138
              I                4.452@
              I
              I F2 - F2         .563*
              I                 .127I
              I                4.446@
              I
              I F3 - F3         .671*
              I                 .117
              I                5.739@
              I                   I
              I F4 - F4        2.316*
              I                 .274
              I                8.443@
              I
                E                D
               ---              ---
E25 -SDQ2N01    1.203*
                 .126
                9.524@
E26 -SDQ2N13    1.123*
                 .125
                9.003@
E27 -SDQ2N25    1.061*
                 .107
                9.882@
E28 -SDQ2N37     .773*
                 .088
                8.797@
E29 -SDQ2N04    1.399*
                 .129
               10.879@
E30 -SDQ2N16     .618*
                 .069
                9.005@
E31 -SDQ2N28     .900*
                 .090
                9.943@
E32 -SDQ2N40     .955*
                 .095
               10.009@
E33 -SDQ2N10     .656*
                 .083
```

TABLE 3.6
(Continued)

```
E34 -SDQ2N22     .660*
                 .076
                8.718@
E35 -SDQ2N34    2.600*
                 .234
               11.108@
E36 -SDQ2N46    1.205*
                 .119
               10.164@
E37 -SDQ2N07     .858*
                 .100
                8.537@
E38 -SDQ2N19    1.232*
                 .122
               10.132@
E39 -SDQ2N31     .366*
                 .065
                5.639@
E40 -SDQ2N43     .967*
                 .093
               10.453@
```

COVARIANCES AMONG INDEPENDENT VARIABLES

```
      V                   F
     ---                 ---
            I F2 - F2      .416*
            I F1 - F1      .079
            I             5.282@
            I
            I F3 - F3      .356*
            I F1 - F1      .072
            I             4.937@
            I
            I F4 - F4      .637*
            I F1 - F1      .119
            I             5.375@
            I
            I F3 - F3      .466*
            I F2 - F2      .079
            I             5.911@
            I
            I F4 - F4      .876*
            I F2 - F2      .135
            I             6.508@
            I
            I F4 - F4      .332*
            I F3 - F3      .101
            I             3.302@
```


TABLE 3.6
(Continued)

```
STANDARDIZED SOLUTION:                                R-SQUARED

SDQ2N01 =V25 =  .581 F1   + .814 E25                    .338
SDQ2N13 =V26 =  .626*F1   + .780 E26                    .391
SDQ2N25 =V27 =  .544*F1   + .839 E27                    .296
SDQ2N37 =V28 =  .640*F1   + .768 E28                    .410
SDQ2N04 =V29 =  .536 F2   + .844 E29                    .287
SDQ2N16 =V30 =  .774*F2   + .634 E30                    .598
SDQ2N28 =V31 =  .702*F2   + .712 E31                    .493
SDQ2N40 =V32 =  .695*F2   + .719 E32                    .483
SDQ2N10 =V33 =  .711 F3   + .703 E33                    .506
SDQ2N22 =V34 =  .668*F3   + .745 E34                    .446
SDQ2N34 =V35 =  .322*F3   + .947 E35                    .104
SDQ2N46 =V36 =  .532*F3   + .847 E36                    .283
SDQ2N07 =V37 =  .854 F4   + .520 E37                    .730
SDQ2N19 =V38 =  .756*F4   + .655 E38                    .571
SDQ2N31 =V39 =  .923*F4   + .385 E39                    .851
SDQ2N43 =V40 =  .712*F4   + .702 E40                    .507
```

CORRELATIONS AMONG INDEPENDENT VARIABLES

```
      V                   F
     ---                 ---
            I F2 - F2      .707*
            I F1 - F1
            I
            I F3 - F3      .555*
            I F1 - F1
            I
            I F4 - F4      .534*
            I F1 - F1
            I
            I F3 - F3      .758*
            I F2 - F2
            I
            I F4 - F4      .767*
            I F2 - F2
            I
            I F4 - F4      .266*
            I F3 - F3
            I
```

and SDQ2N07 (V37), all information appears on one line only—these represent the fixed factor-loading parameters; therefore, no estimated values are presented. For each of the estimated (*) parameters, however, there are three lines of output: the estimated value is presented first, the standard error second, and the test statistic last, with statistically significant parameters assigned an @. Review of the unstandardized solution in Table 3.6 shows all estimates to be reasonable as well as statistically significant; all standard errors also appear to be in good order.

In the standardized solution, all variables are rescaled to have a variance of 1.0. In EQS, standardization is applied to all variables in the linear structural equation system, including errors and disturbances. As a result, all coefficients in the equations have a similar interpretation and the magnitude of their standardized values may be easier to interpret than that of coefficients obtained from the covariance or raw data metric (Bentler, 2005).[12] In contrast to the unstandardized solution, information related to the standardized solution is summarized on one line, along with a related R^2 value (i.e., the squared multiple correlation) appearing in the column to the right labeled R-SQUARED (see Aiken, West, & Pitts, 2003, p. 485). Each measured (or dependent, in the Bentler–Weeks sense) variable is accompanied by an R^2 value representing the proportion of variance accounted for by its related factor (or independent predictor variable). It is computed by subtracting the square of the error term from 1.0 (i.e., $1.0 - E^2$). Bentler (2005) notes that in the event that a particular R^2 cannot be computed or that it differs by more than 0.01 from the corresponding Bentler–Raykov corrected R^2 value (Bentler & Raykov, 2000), the corrected R^2 value will be printed below.

In reviewing the standardized estimates, the user should verify that particular parameter values are consistent with the literature. For example, within the context of the present application, it is of interest to inspect correlations among the SC factors for their consistency with previously reported values; in the present example, these estimates are as expected.

Three additional features of the standardized solution are note worthy. First, in the event that some variances are estimated as negative values, the standardized solution cannot be obtained because the computation requires the square roots of these values; if such is the case, no standardized solution will be printed. Second, in the standardized solution, parameters that were previously fixed to 1.0 take on new values. Finally, note the absence of output for the variances of the independent variables. This is because in standardizing the estimates, the variances automatically take on a value of 1.00.

Model Misspecification. Determination of misfitting parameters is accomplished in EQS by means of the LM Test. As discussed previously, fixed parameters—as specified in the input file—are assessed both univariately and multivariately to identify parameters that would contribute to a significant drop in χ^2 if they were to be freely estimated in a subsequent EQS run. More specifically, information is presented first for univariate LM Tests of model parameters constrained either to 0.0 or to some nonzero value. If any of the univariate tests yield statistically significant results, the program then proceeds with a

[12]The standardized solution in EQS is not the same as the one in the LISREL output. In the latter, neither the measured variables, error terms, nor disturbances in equations (not present in first-order CFA models) are standardized (Bentler, 2005).

multivariate test of fixed parameters. As such, it proceeds with a forward stepwise procedure that selects as the next parameter to be added to the multivariate test, the single fixed parameter that provides the largest increase in the multivariate χ^2 statistic (Bentler, 2005). Results related to these LM Tests are presented in Table 3.7.

Univariate Test Statistics

Review of these results shows eight columns of information. Column 1 simply assigns a number to the parameter being tested. Column 2 assigns a dual numerical code to the parameter under test. In a simultaneous test of parameters, which is the case here, the first digit will always be 2, as shown in Table 3.7. The second digit refers to the submatrix number in which the parameter resides (1 to 22). However, if the parameter has been fixed to a nonzero value, the second digit will be zero. (There is no reason to remember this information; it is presented solely in the interest of completeness.) Column 3 lists the parameter under test.

Column 4 presents the univariate LM χ^2 test statistic, which has 1 degree of freedom; Column 5 presents its related statistical probability. These values result from testing the hypothesized constraint that the selected parameter is equal to zero. Interpreted literally, given a probability of less than .05, this hypothesis must be rejected, thereby indicating some evidence of misspecification in the model. However, Bentler (2005) cautions that because these LM univariate tests are correlated and can be applied repeatedly to test a variety of single restrictions, they should not be used to determine what the simultaneous effect of several restrictions may be. Such decisions should always be based on the multivariate LM Test results.

Column 6, labeled Hancock 98 df Prob presents LM Test probabilities based on Hancock's (1999) multiple comparison rationale. This criterion was developed for use in SEM as an analog to the Scheffé test (1953) used in ANOVA to control for family-wise Type I errors. These probabilities represent an evaluation of each LM χ^2 statistic based on degrees of freedom for the current model (98 in the present case) rather than the usual 1 degree of freedom. Hancock's criterion provides for an extremely conservative approach (too conservative, in my experience) to model testing that can help control Type I error in exploratory (i.e., post hoc) model modification.

Columns 7 and 8 present the unstandardized and standardized parameter change statistic, respectively.[13] For each parameter tested via the LM Test, the parameter change statistic represents its estimated value if this parameter is freely estimated in a subsequent test of the model. Such information can be helpful in determining whether a parameter identified by the LM Test is justified to stand as a candidate for respecification as a freely estimated parameter. In other words, is the estimated

[13]This is sometimes referred to as the expected parameter change statistic.