Chapter 5
Structural Equation Models with Latent Variables

Structural Equation Modeling of Şenol-Durak and Ayvaşik's Posttraumatic Growth Data

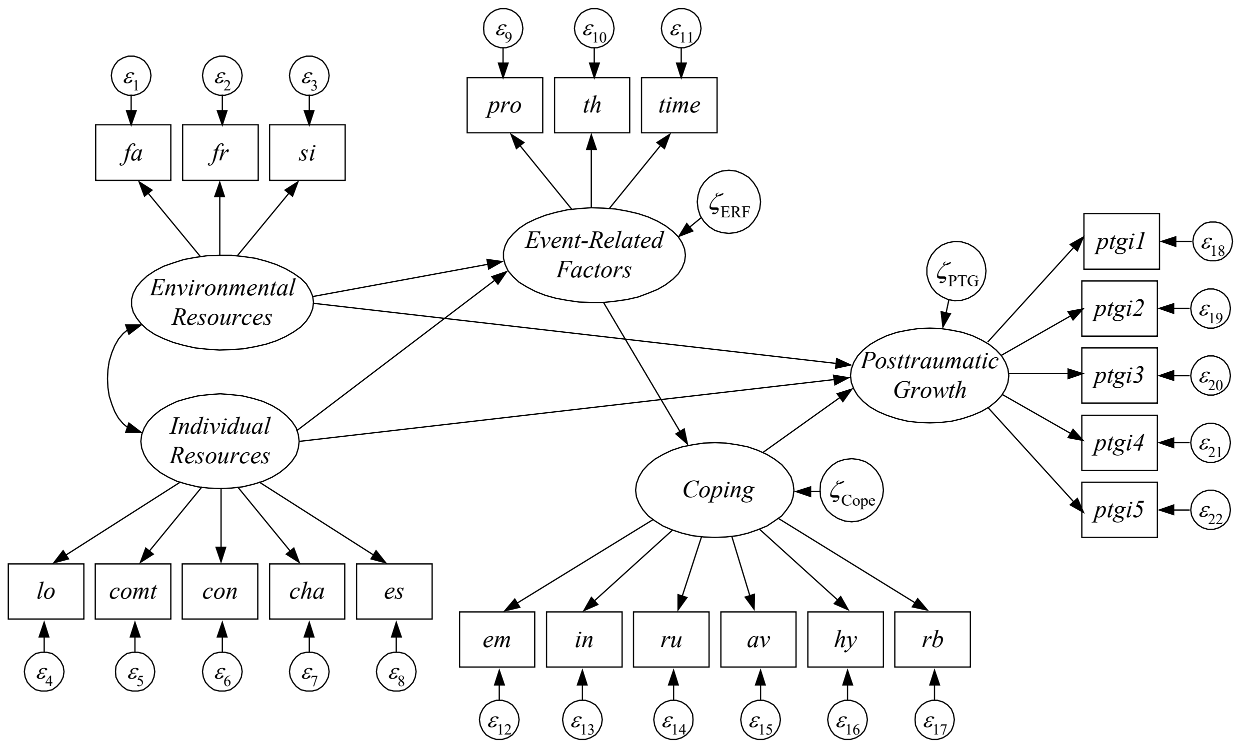
The data for this demonstration were provided by Şenol-Durak & Ayvaşik in their 2010 *Journal of Health Psychology* manuscript, "Factors associated with posttraumatic growth among the spouses of myocardial infraction patients." The sample includes 132 spouses of myocardial infarction patients. The correlation matrix as well as the means and standard deviations for the variables were provided by the authors. This information is in the text file **`mip.dat`**. The variables in the data set that we will use are

`fa`	social support from family	*Environmental Resources*
`fr`	social support from friends	
`si`	social support from significant others	
`lo`	Locus of Control Scale score	*Individual Resources*
`comt`	Commitment score	
`con`	Control score	
`cha`	Challenge score	
`es`	Rosenberg Self Esteem Scale score	
`pro`	subjective evaluation of prognosis	*Event Related Factors*
`th`	threat to future health	
`time`	time since diagnosis	
`em`	emotion focused coping	*Cognitive Process Coping*
`in`	indirect coping	
`ru`	rumination	
`av`	avoidance	
`hy`	hypervigilance	
`rb`	religious beliefs	
`ptgi1`	improved relationships	*Posttraumatic Growth*
`ptgi2`	new possibilities for one's life	
`ptgi3`	greater appreciation of life	
`ptgi4`	greater sense of personal strength	
`ptgi5`	spiritual development	

Refer to the article for definitions of variables not included in the model.

Initial Hypothesized Model

The hypothesized model for the data predicts that both individual and environmental resources directly lead to increased posttraumatic growth, but also indirectly lead to somewhat decreased posttraumatic growth by reducing event-related hardship, thus decreasing the need for coping and reducing opportunities for posttraumatic growth. The hypothesized model also predicts that neither environmental nor individual resources have a direct impact on cognitive coping. Further, the effect of event-related factors on posttraumatic growth is hypothesized to be purely mediated by coping.



We can also express the model using matrix algebra, as shown on the next page.

The measurement model is:

$$\begin{bmatrix} fa_i \\ fr_i \\ si_i \\ lo_i \\ comt_i \\ con_i \\ cha_i \\ es_i \\ pro_i \\ th_i \\ time_i \\ em_i \\ in_i \\ ru_i \\ av_i \\ hy_i \\ rb_i \\ ptgi_{1i} \\ ptgi_{2i} \\ ptgi_{3i} \\ ptgi_{4i} \\ ptgi_{5i} \end{bmatrix} = \begin{bmatrix} \nu_{1i} \\ \nu_{2i} \\ \nu_{3i} \\ \nu_{4i} \\ \nu_{5i} \\ \nu_{6i} \\ \nu_{7i} \\ \nu_{8i} \\ \nu_{9i} \\ \nu_{10i} \\ \nu_{11i} \\ \nu_{12i} \\ \nu_{13i} \\ \nu_{14i} \\ \nu_{15i} \\ \nu_{16i} \\ \nu_{17i} \\ \nu_{18i} \\ \nu_{19i} \\ \nu_{20i} \\ \nu_{21i} \\ \nu_{22i} \end{bmatrix} + \begin{bmatrix} \lambda_{11} & 0 & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 & 0 \\ \lambda_{31} & 0 & 0 & 0 & 0 \\ 0 & \lambda_{42} & 0 & 0 & 0 \\ 0 & \lambda_{52} & 0 & 0 & 0 \\ 0 & \lambda_{62} & 0 & 0 & 0 \\ 0 & \lambda_{72} & 0 & 0 & 0 \\ 0 & \lambda_{82} & 0 & 0 & 0 \\ 0 & 0 & \lambda_{93} & 0 & 0 \\ 0 & 0 & \lambda_{10,3} & 0 & 0 \\ 0 & 0 & \lambda_{11,3} & 0 & 0 \\ 0 & 0 & 0 & \lambda_{12,4} & 0 \\ 0 & 0 & 0 & \lambda_{13,4} & 0 \\ 0 & 0 & 0 & \lambda_{14,4} & 0 \\ 0 & 0 & 0 & \lambda_{15,4} & 0 \\ 0 & 0 & 0 & \lambda_{16,4} & 0 \\ 0 & 0 & 0 & \lambda_{17,4} & 0 \\ 0 & 0 & 0 & 0 & \lambda_{18,5} \\ 0 & 0 & 0 & 0 & \lambda_{19,5} \\ 0 & 0 & 0 & 0 & \lambda_{20,5} \\ 0 & 0 & 0 & 0 & \lambda_{21,5} \\ 0 & 0 & 0 & 0 & \lambda_{22,5} \end{bmatrix} \begin{bmatrix} \eta_{ERi} \\ \eta_{IRi} \\ \eta_{ERFi} \\ \eta_{COPE_i} \\ \eta_{PTGi} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \\ \varepsilon_{5i} \\ \varepsilon_{6i} \\ \varepsilon_{7i} \\ \varepsilon_{8i} \\ \varepsilon_{9i} \\ \varepsilon_{10i} \\ \varepsilon_{11i} \\ \varepsilon_{12i} \\ \varepsilon_{13i} \\ \varepsilon_{14i} \\ \varepsilon_{15i} \\ \varepsilon_{16i} \\ \varepsilon_{17i} \\ \varepsilon_{18i} \\ \varepsilon_{19i} \\ \varepsilon_{20i} \\ \varepsilon_{21i} \\ \varepsilon_{22i} \end{bmatrix}$$

where $\mathbf{\Theta} = DIAG\left(\theta_{11}, \theta_{22}, \ldots, \theta_{22,22}\right)$

The latent variable model is:

$$\begin{bmatrix} \eta_{ERi} \\ \eta_{IRi} \\ \eta_{ERFi} \\ \eta_{COPE_i} \\ \eta_{PTGi} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 & 0 & 0 \\ 0 & 0 & \beta_{43} & 0 & 0 \\ \beta_{51} & \beta_{52} & 0 & \beta_{54} & 0 \end{bmatrix} \begin{bmatrix} \eta_{ERi} \\ \eta_{IRi} \\ \eta_{ERFi} \\ \eta_{COPE_i} \\ \eta_{PTGi} \end{bmatrix} + \begin{bmatrix} \zeta_{ERi} \\ \zeta_{IRi} \\ \zeta_{ERFi} \\ \zeta_{COPE_i} \\ \zeta_{PTGi} \end{bmatrix}$$

where $\mathbf{\Psi} = \begin{bmatrix} 1 & & & & \\ \psi_{21} & 1 & & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & 1 & \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

Note that the means/intercepts and (residual) variances of the factors have been fixed to 0 and 1, respectively to scale the latent variables. In class, we used the *t*-rule and the two-step rule to verify that the model is identified. We can thus go on to specify the model in Mplus.

The Mplus input file that fits this model is provided in **`ch05_1.inp`** and is shown below:

```
TITLE:
  Senol-Durak & Ayvasik SEM;

DATA:
  FILE IS mip.dat;
  TYPE IS FULLCORR MEANS STDEVIATIONS;
  NOBSERVATIONS=132;

VARIABLE:
  NAMES ARE ptgi ptgi1 ptgi2 ptgi3 ptgi4 ptgi5 marital fa fr si child
   child18 age gender depres comt con cha es lo pro th diord time
   problem em in ru av hy relipart rb;

  USEVARIABLES ARE fa fr si lo comt con cha es
            pro th time em in ru av hy rb
            ptgi1 ptgi2 ptgi3 ptgi4 ptgi5;

ANALYSIS:
  ESTIMATOR=ML;

MODEL:
  ER by fa* fr si;
  IR by lo*-1 comt con cha es;
  ERF by pro*1 th*-1 time*1;
  CPP by em*1 in*-1 ru av hy rb;
  PTG by ptgi1* ptgi2 ptgi3 ptgi4 ptgi5;
  [fa fr si];
  [lo comt con cha es];
  [pro th time];
  [em in ru av hy rb];
  [ptgi1 ptgi2 ptgi3 ptgi4 ptgi5];
  [ER@0 IR@0 ERF@0 CPP@0 PTG@0];
  ER@1 IR@1 ERF@1 CPP@1 PTG@1;
  ER with IR;
  ERF on ER IR;
  CPP on ERF;
  PTG on ER IR CPP;

OUTPUT:
  sampstat stdyx mod;
```

We have seen most of these commands before, so here we will highlight only portions of the code.

Mplus accepts either raw data or summary data (i.e., means, standard deviations, and correlations among variables). As a default, Mplus assumes that data are in raw format. Since the data from this example are in summary form, we used the `DATA` command to tell the program that we are inputting a correlation matrix by writing: `TYPE IS FULLCORR MEANS STDEVIATIONS`. We then included the sample size with `NOBSERVATIONS`.

As before, the measurement models are specified under `MODEL` using the `by` statements. Here, we have used asterisks to allow all factor loadings to be freely estimated. Mplus defaults to fixing the first factor loading to 1, so it is only necessary to place an asterisk next to the first loading on each factor in order to ensure that all factor loadings are freely estimated.

In the journal article, some of the estimated factor loadings were negative. Due to rotational indeterminacy, a measurement model will fit equally well if all of its factor loadings are directionally flipped such that positive loadings are negative and negative loadings are positive. We provided starting values to ensure the model would converge to the most interpretable solution. Thus, for example, event related factors (**`ERF`**) was given positive starting values for **`pro`** and **`time`** and a negative starting value for **`th`** so that this factor has a positive valence (higher values are better, e.g., indicating perception of better prognosis and less life threat). This was achieved by placing a starting value after an asterisk (e.g., 1 or -1).

We have standardized the factors in order to identify the measurement models. Factor means (and intercepts for endogenous factors) were fixed to zero by placing `@0` next to each factor name inside of square brackets. In Mplus, square brackets denote means and intercepts and `@` is used to constrain parameters to a fixed value. Factor variances (and residual variances for endogenous factors) were constrained to 1 by placing an `@1` next to each factor name on a line without brackets.

Structural covariances are specified in Mplus using the `with` statement, and regression parameters are specified using the `on` statement (outcome variable `on` predictor variable).

We have requested sample statistics (`sampstat`), the typical standardized solution (with both the predictor and outcome standardizes; `stdyx`), and modification indices (`mod`) under the `OUTPUT` command.

Let us now turn to the fit indices for the model:

```
Number of Free Parameters                       73

Loglikelihood

          H0 Value                       -8012.004
          H1 Value                       -7837.004

Information Criteria

          Akaike (AIC)                   16170.009
          Bayesian (BIC)                 16380.453
          Sample-Size Adjusted BIC       16149.553
            (n* = (n + 2) / 24)
```

```
Chi-Square Test of Model Fit

          Value                           350.000
          Degrees of Freedom                  202
          P-Value                          0.0000

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                          0.075
          90 Percent C.I.                   0.061  0.087
          Probability RMSEA <= .05          0.002

CFI/TLI

          CFI                               0.848
          TLI                               0.826

Chi-Square Test of Model Fit for the Baseline Model

          Value                          1205.231
          Degrees of Freedom                  231
          P-Value                          0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                             0.101
```

The fit indices indicate that the model does not fit the data well. Rather than interpreting the parameter estimates, we will examine modification indices to get a sense for what might be causing the model to fit poorly, keeping in mind that any model modification must be theoretically justifiable.

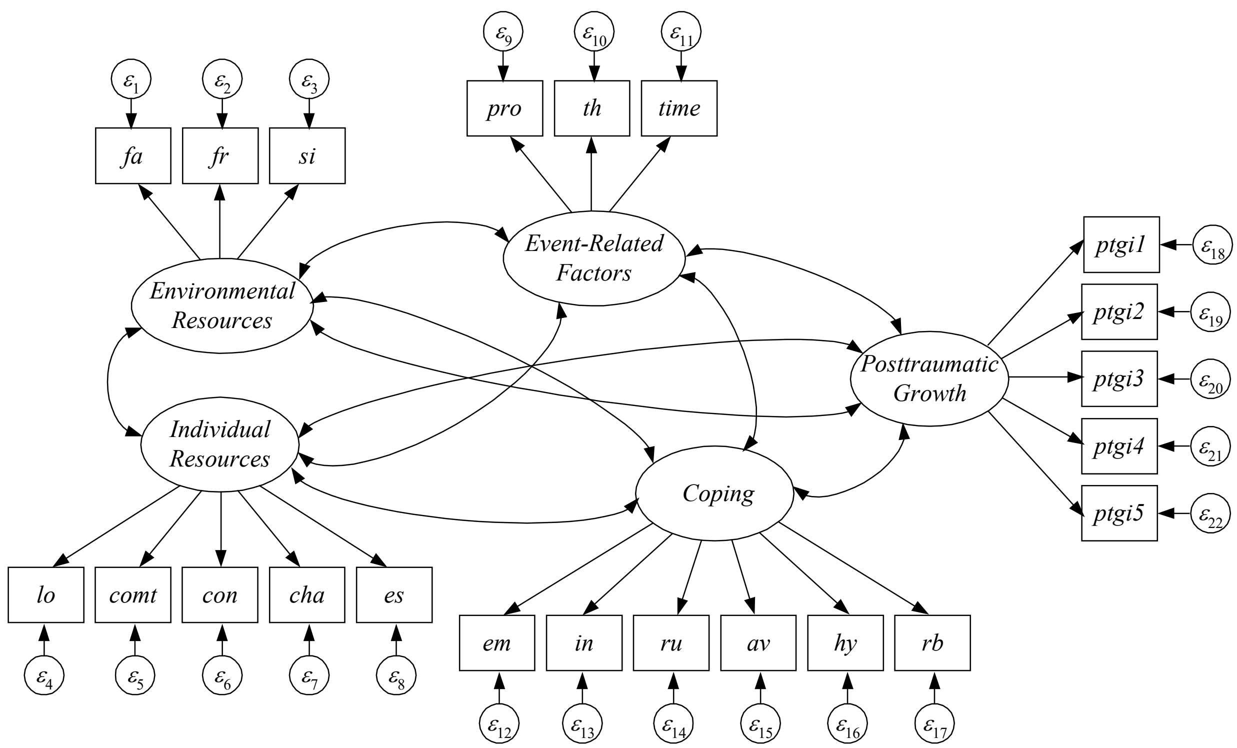
```
Minimum M.I. value for printing the modification index    10.000

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

BY Statements

IR       BY PTGI5          11.877    -0.730     -0.730        -0.241
PTG      BY CHA            10.321     0.590      0.688         0.282

WITH Statements

CHA      WITH COMT         27.844     2.684      2.684         0.546
CHA      WITH CON          12.343    -2.215     -2.215        -0.381
EM       WITH CON          10.695    -8.717     -8.717        -0.304
IN       WITH FA           14.250    -7.574     -7.574        -0.333
IN       WITH EM           21.340   -26.655    -26.655        -0.410
HY       WITH RU           32.248    34.384     34.384         4.461
```

Modification indices suggest that the largest improvement to the model chi-square could be achieved by allowing hypervigilance to correlate with rumination, over and above the correlation implied by the coping factor, allowing challenge to correlate with commitment over and above the individual resources factor, and allowing indirect and emotional coping to correlate above and beyond the correlation implied by the coping factor. These modifications reflect misspecification in the measurement model.

Confirmatory Factor Analysis

When building a structural equation model, a useful strategy to avoid complex misspecification is to begin by ensuring that the simplest foundation of the overall model, the measurement model, is correctly specified. Once the measurement model has been properly specified, the next step is to incorporate structural parameters. Thus, we turn next to a CFA with saturated covariances among factors. This strategy will allow us to get measurement right so that measurement misspecification is not confounded with structural misfit.

The CFA model is provided in **`ch05_2.inp`**, shown below.



We omit discussion of the input file because CFA estimation was discussed in the previous chapter. The resulting model fit is shown below.

```
MODEL FIT INFORMATION

Number of Free Parameters                       76

Loglikelihood

          H0 Value                       -8011.322
          H1 Value                       -7837.004

Information Criteria

          Akaike (AIC)                   16174.645
          Bayesian (BIC)                 16393.738
          Sample-Size Adjusted BIC       16153.348
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                            348.636
          Degrees of Freedom                   199
          P-Value                           0.0000

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.075
          90 Percent C.I.                    0.062  0.088
          Probability RMSEA <= .05           0.001

CFI/TLI

          CFI                                0.846
          TLI                                0.822

Chi-Square Test of Model Fit for the Baseline Model

          Value                           1205.231
          Degrees of Freedom                   231
          P-Value                           0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                              0.099
```

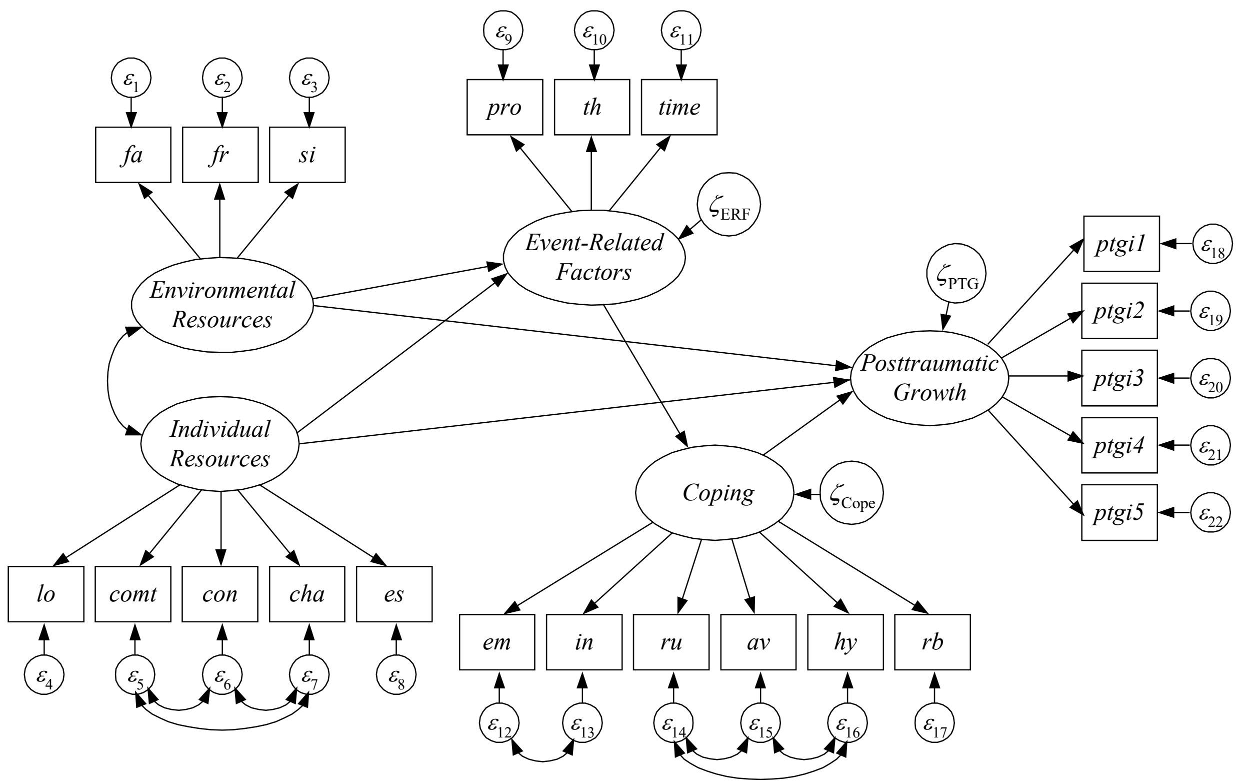
The model still does not fit the data well, confirming our hypothesis that the measurement model, and not the structural model, is misspecified. Indeed, we can conduct a likelihood ratio test comparing the CFA model with the hypothesized model because the hypothesized model is a constrained version of the CFA with three structural parameters fixed to zero:

$$\Delta\chi^2(3) = 350.00 - 348.64 = 1.36, p = .71$$

Combining this information with the information provided earlier from the modification indices, we can conclude that the measurement model requires respecification. Theoretically, allowing some residuals to correlate (as suggested by the MIs) makes sense because some factors include multiple subscale scores as indicators. When combined with other items from different scales, we would expect some degree of local dependence. Specifically, the **IR** factor includes three Psychological Hardiness subscale scores as indicators (**comt**, **con**, and **cha**), but also two indicators from independent scales (**lo** and **es**). The coping factor includes two indicators from the Ways of Coping Inventory (**em** and **in**), three indicators from the Impact of Event Scale (**ru**, **av**, and **hy**), and a religious beliefs score from another scale (**rb**).

Revised Model

We now introduce correlated uniquenesses for **comt**, **con**, and **cha** on the individual resources factor, between **em** and **in** on the coping factor, and among **ru**, **av**, and **hy** on the coping factor.



The new correlated uniquenesses can be included by modifying the MODEL syntax in Mplus. The Mplus input file that fits this model is provided in **ch05_3.inp**.

```
MODEL:
  ER by fa* fr si;
  IR by lo*-1 comt con cha es;
  ERF by pro*1 th*-1 time*;
  CPP by em*1 in*-1 ru* av* hy* rb*;
  PTG by ptgi1* ptgi2 ptgi3 ptgi4 ptgi5;
  [fa fr si];
  [lo comt con cha es];
  [pro th time];
  [em in ru av hy rb];
  [ptgi1 ptgi2 ptgi3 ptgi4 ptgi5];
  [ER@0 IR@0 ERF@0 CPP@0 PTG@0];
  ER@1 IR@1 ERF@1 CPP@1 PTG@1;
  ER with IR;
  ERF on ER IR;
  CPP on ERF;
  PTG on ER IR CPP;
  comt with con cha;
  con with cha;
  em with in;
  ru with av hy;
  av with hy;
```

Mplus does not use separate names for uniquenesses/residuals/disturbances. Instead, uniquenesses or disturbances are referred to by the referent variable. Thus, covariances between uniquenesses are specified via the `with` statement just as covariances among variables are. The line `con with cha` thus includes a covariance between the uniquenesses of **con** and **cha**.

We use the `MODEL INDIRECT` command to request calculation of the total indirect effects of one variable on another. The `IND` statement is used to request an estimate of the indirect effect of a predictor on an outcome by listing the outcome first, followed by `IND`, and then the predictor. Here, we have requested indirect effect estimates of **ERF**, **IR**, and **ER** on **PTG**. (Note we are not using the bootstrap procedures described in Chapter 3 because that procedure requires raw data and here we are fitting the model to the summary statistics.)

```
MODEL INDIRECT:
  PTG IND ERF;
  PTG IND IR;
  PTG IND ER;
```

The resulting output is shown here:

```
MODEL FIT INFORMATION

Number of Free Parameters                        80

Loglikelihood

          H0 Value                        -7972.143
          H1 Value                        -7837.004
```

```
Information Criteria

          Akaike (AIC)                    16104.287
          Bayesian (BIC)                  16334.911
          Sample-Size Adjusted BIC        16081.869
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                             270.278
          Degrees of Freedom                    195
          P-Value                            0.0003

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                            0.054
          90 Percent C.I.                     0.037  0.069
          Probability RMSEA <= .05            0.325

CFI/TLI

          CFI                                 0.923
          TLI                                 0.908

Chi-Square Test of Model Fit for the Baseline Model

          Value                            1205.231
          Degrees of Freedom                    231
          P-Value                            0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                               0.090
```

Adding 7 free parameters to the hypothesized model resulted in a significant improvement in model fit:

$\Delta\chi^2(7) = 350.00 - 270.28 = 79.72, p < .001$.

Other fit indices suggest that the modified model has a satisfactory fit to the data.

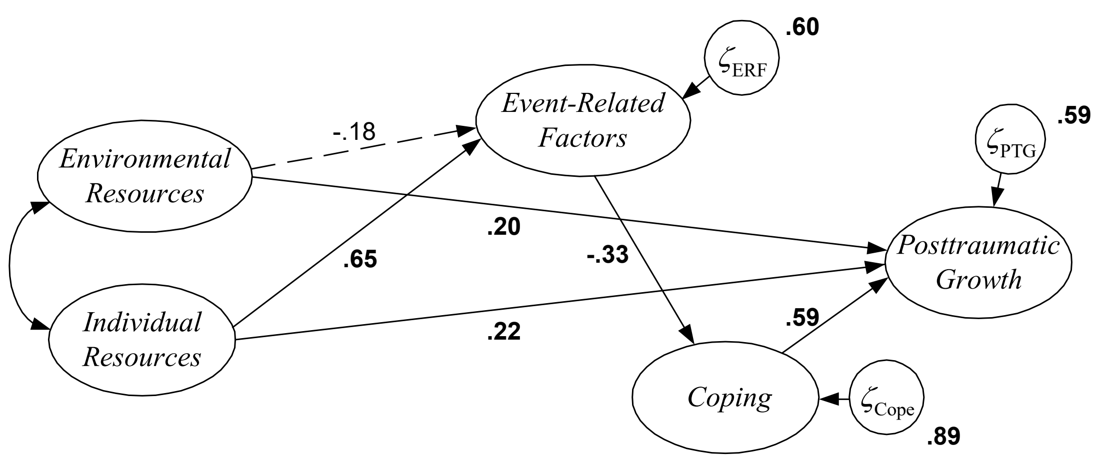
Parameter estimates are presented below.

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ER	BY				
FA		1.317	0.362	3.637	0.000
FR		5.855	0.767	7.633	0.000
SI		4.457	0.776	5.740	0.000
IR	BY				
LO		-11.555	1.675	-6.897	0.000
COMT		1.028	0.290	3.538	0.000
CON		2.312	0.325	7.105	0.000
CHA		1.141	0.264	4.324	0.000
ES		3.007	0.551	5.453	0.000
ERF	BY				
PRO		0.347	0.103	3.365	0.001
TH		-0.342	0.114	-3.012	0.003
TIME		1.316	0.743	1.771	0.076
CPP	BY				
EM		6.060	1.103	5.495	0.000
IN		-3.100	0.668	-4.644	0.000
RU		4.173	0.798	5.229	0.000
AV		2.958	0.587	5.039	0.000
HY		3.725	0.625	5.963	0.000
RB		0.221	0.088	2.515	0.012
PTG	BY				
PTGI1		6.036	0.633	9.538	0.000
PTGI2		4.578	0.484	9.457	0.000
PTGI3		2.974	0.365	8.146	0.000
PTGI4		2.259	0.261	8.649	0.000
PTGI5		1.879	0.214	8.794	0.000
ERF	ON				
ER		-0.236	0.181	-1.305	0.192
IR		0.834	0.293	2.840	0.005
CPP	ON				
ERF		-0.276	0.130	-2.120	0.034
PTG	ON				
ER		0.259	0.118	2.201	0.028
IR		0.284	0.139	2.039	0.041
CPP		0.722	0.168	4.299	0.000

IR has a significant direct effect on **ERF** ($\gamma = .834$; S.E. = .293; *p* = .005)and **PTG** ($\gamma = .284$; S.E. = .139; *p* = .041). **ER** has a significant direct effect on **PTG** ($\gamma = .259$; S.E. = .118; *p* = .28) and **coping** ($\gamma = -.276$; S.E. = .130; *p* = .034). **Coping** is significantly related to **PTG** ($\gamma = .722$; S.E. = .168 *p* < .001). Next we can view the standardized parameter estimates.

```
STDYX Standardization

ERF      ON
    ER                  -0.183      0.131     -1.404      0.160
    IR                   0.648      0.139      4.671      0.000

 CPP      ON
    ERF                 -0.334      0.135     -2.478      0.013

 PTG      ON
    ER                   0.200      0.087      2.286      0.022
    IR                   0.219      0.100      2.192      0.028
    CPP                  0.590      0.091      6.497      0.000
Variances
    ER                   1.000      0.000    999.000    999.000
    IR                   1.000      0.000    999.000    999.000

 Residual Variances
    ERF                  0.604      0.168      3.602      0.000
    CPP                  0.888      0.090      9.859      0.000
    PTG                  0.593      0.102      5.795      0.000
```

We focus on the structural parameter estimates in this chapter because interpretation of measurement models has been discussed previously. The standardized structural parameter estimates have been drawn on the path diagram below to more easily comprehend the model results. The non-significant path from environmental resources to event-related factors is dashed. All other paths are statistically significant and shown with solid lines.



Standardized results suggest that environmental and individual resources have a moderate, direct, positive influence on posttraumatic growth, cognitive coping has a strong, direct, positive influence on posttraumatic growth, individual resources strongly predict more event-related factors (shorter time since prognosis, poorer prognosis, and greater threat), and more positive event-related factors predicts moderately less cognitive coping. Individual resources and event related factors have a complex relationship with posttraumatic growth. To better understand these relationships, we must consider direct, indirect, and total effects.

```
TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

                                                     Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

Effects from ERF to PTG

  Total               -0.199      0.100     -1.982      0.047
  Total indirect      -0.199      0.100     -1.982      0.047

  Specific indirect

    PTG
    CPP
    ERF               -0.199      0.100     -1.982      0.047


Effects from IR to PTG

  Total                0.118      0.132      0.893      0.372
  Total indirect      -0.166      0.090     -1.833      0.067

  Specific indirect

    PTG
    CPP
    ERF
    IR                -0.166      0.090     -1.833      0.067

  Direct
    PTG
    IR                 0.284      0.139      2.039      0.041


Effects from ER to PTG

  Total                0.306      0.123      2.494      0.013
  Total indirect       0.047      0.041      1.152      0.249

  Specific indirect

    PTG
    CPP
    ERF
    ER                 0.047      0.041      1.152      0.249

  Direct
    PTG
    ER                 0.259      0.118      2.201      0.028
```

For each predictor-to-outcome effect, the Mplus output first presents the total effect (along with standard errors and significance tests). Then it breaks down the total effect by presenting the total indirect effect and the direct effect. In some cases, the entire effect is indirect (e.g., `ERF to PTG`). If the indirect effect consists of multiple pathways, the indirect effect is further divided to show the specific indirect effect for each pathway. In this example, each predictor only had one indirect pathways affecting **PTG**.

We will closely examine the effect of **IR** on **PTG**. We start by noting that the total effect of **IR** on **PTG** is non-significant. However, upon closer examination, it is apparent that **IR** is related to **PTG** both directly and indirectly, but that these effects are in opposite directions such that the net, total effect is nearly zero. The direct effect of **IR** on **PTG** is significant and positive, but the indirect effect is marginally significant and negative.