Confirmatory Factor Analysis Using M*plus*

In this document I explain how to use M*plus* to run confirmatory factor analyses.

The data from this study are responses of 1022 people to items on the Achievement Goal Questionnaire (Elliot & McGregor, 2001) collected as part of a study by Finney, Pieper, and Barron (2004) who were interested in determining whether the factor structure reported by Elliott and McGregor could be replicated in a general, rather than a specific, academic context. The response scale ranged from 1 = not at all true of me to 7 = very true of me. There are no missing data, but I have included the "missing" command used to specify missing data in *Mplus* to demonstrate how this would be done. The data are in tab-delimited format and can be found in the file "goal orientation.dat"

CFA Syntax

Below I present the M*plus* syntax, followed by an explanation of each command. M*plus* commands are in boldface. Non-boldface words indicate information that must be provided by the researcher for their specific data.

```
Title: CFA of goal orientation data;
data: file is goal2.dat;
        format is free;
variable: names are i1 i2 i3 i4 i5 i6 i7 i8 i9 i10 i11 i12;
            missing all (8,9,10);
model:
          perfapp by i1 i2 i3;
          peravoid by i4 i5 i6;
          masavoid by i7 i8 i9;
          masapp by i10 i11 i12;
          [i1 – i12];
output:  sampstat residual stand(stdyx) modindices;
```

Each M*plus* command begins with a keyword followed by a colon. Each command must end with a semi-colon.

The **title** command specifies a title that will be printed at the top of each page of output. It is not required, but should be included for easy identification of different sets of output.

The **data** command is required and specifies the name and location of the data file. In this example I do not include a directory specification for the file because the data are in the same location as the M*plus* program. If this is not the case, the name of the data file should be preceded by the directory in which the file resides.

Data in the goal orientation file are tab-delimited. Data that are tab-delimited or in which variables are separated by at least one space do not require a format statement. I include the statement

format is free, although this is the default specification, so if data are separate by spaces or tabs, this statement is not required.

The **variables** command specifies the names of the variables in the data file. Variable names should be eight characters or fewer. The variable names must be in the same order as the variables appear in the data file. The **missing** subcommand specifies that values of 0,8,and 9 are codes for missing values for all the variables (indicated by the keyword **all**). Inclusion of the **missing** subcommand automatically invokes full information maximum likelihood (FIML) missing data handling for CFA models.

The **model** command specifies the model to be analyzed. The keyword ***by*** means "measured by" and is used to specify a confirmatory factor analysis model in which four latent variables ("perfapp," "peravoid," "masavoid," and "masapp") are each measured by three observed variables. Here, the names of the latent variables are specified by the researcher and must follow the same length conventions as other variables. The names of the observed variables come after the keyword **by**.

The specification **[i1-i12]** indicates that the mean, or intercept, of each observed variable (i1 - i12) should be estimated.

By default, latent factors in CFAs are correlated and factor variances as well as each variable's residual variance are estimated.

Also by default, M*plus* sets the first loading for each factor to 1.0 to identify the model (in our example, the loadings for i1, i4, i7, and i10 would be set to 1.0).

In some cases, researchers prefer to identify the model (see Chapter 13 for more information on identification) by setting the factor variances to 1.0. To do this, replace the **model** syntax with the syntax below:

```
perfapp by i1* i2 i3;
peravoid by i4* i5 i6;
masavoid by i7* i8 i9;
masapp by i10* i11 i12;
perfapp@1 peravoid@1 masavoid@1 masapp@1;
```

The asterisks after i1, i4, i7, and i10 indicate that the loadings for these variables should be estimated rather than fixed to 1.0

The last command sets the variances of the four factors to 1.0 with the specification "@1."

Finally, the **output** line requests the sample statistics (**sampstat**), the model residuals (**residual**), the standardized solution (**stand(stdyx)**), and modification indices (**modindices**).

The sample statistics include the variable means, variances, skewness and kurtosis values of the observed variables as well as their covariance and correlation matrices.

The model residuals are the differences between the observed and model-implied covariances. M*plus* provides several different standardized solutions. The specification **stdyx** requests the completely standardized solution, in which both the latent and observed variables are standardized.

The modification indices are estimates of the amount by which the chi-square value would decrease if the associated parameter were to be included in the model.

See pages 383-388 in the text for more information on residuals and modification indices.

CFA Output

In this section I present selected output from the analysis.

Information on model fit is presented first. Here I show only the indices discussed in the text (see pages 376-381 for more information).

```
MODEL FIT INFORMATION

Chi-Square Test of Model Fit

          Value                            283.977
          Degrees of Freedom                48
          P-Value                          0.0000

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.069
          90 Percent C.I.                    0.062  0.077
          Probability RMSEA <= .05           0.000

CFI/TLI

          CFI                                0.946
          TLI                                0.926

SRMR (Standardized Root Mean Square Residual)

          Value                              0.046
```

These values indicate some model misspecification.

The chi-square is a test of whether the observed covariance matrix is equal to the model-implied matrix, or, more broadly, whether the model fits the data. Rejection of the chi-square test indicates that the model does not fit the data.

A common cut-off value for the RMSEA is .06, with higher values indicating an unacceptable lack of fit.
The CFI compares the fit of the model to that of a null, or baseline model, and is a measure of the degree to which the model fits better than a model in which the variables are mutually uncorrelated. A common cut-off value is .95.

Finally, the SRMR is a function of the average of the residuals (observed – model-implied). Common cut-off values are .08-.09.

The model parameter estimates are presented next. Here, I present the unstandardized results. The completely standardized are shown in Figure 13.4 in the text (p. 373).

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
PERFAPP BY				
I1	1.000	0.000	999.000	999.000
I2	1.035	0.036	28.865	0.000
I3	1.052	0.036	29.396	0.000
PERAVOID BY				
I4	1.000	0.000	999.000	999.000
I5	0.772	0.060	12.807	0.000
I6	1.195	0.081	14.837	0.000
MASAVOID BY				
I7	1.000	0.000	999.000	999.000
I8	1.956	0.150	13.001	0.000
I9	1.845	0.139	13.270	0.000
MASAPP BY				
I10	1.000	0.000	999.000	999.000
I11	0.974	0.053	18.472	0.000
I12	1.041	0.058	18.031	0.000

The values in the “estimate” column above are the factor loadings or coefficients. These should be large in magnitude and in the expected direction.

The standard errors are shown in the next column (labeled “S.E.).

The third column contains the *z*-values, which test the hypothesis that the parameter estimate is equal to zero.

Finally, the two-tailed *p*-value for the *z*-tests are shown in the last column.

PERAVOID WITH				
PERFAPP	0.751	0.071	10.612	0.000
MASAVOID WITH				
PERFAPP	0.197	0.034	5.837	0.000
PERAVOID	0.378	0.046	8.171	0.000
MASAPP WITH				
PERFAPP	0.308	0.043	7.234	0.000
PERAVOID	0.080	0.039	2.063	0.039
MASAVOID	0.159	0.025	6.477	0.000

The values above are the covariances among the four factors. The standardized solution provides the factor correlations.

Variances				
PERFAPP	1.575	0.105	14.968	0.000
PERAVOID	1.310	0.148	8.857	0.000
MASAVOID	0.432	0.062	7.010	0.000
MASAPP	0.663	0.060	11.037	0.000

The values above are the variances of the latent variables, or factors.

Below these are the residual, or unexplained, variance in the observed variables.

Residual Variances				
I1	0.816	0.050	16.358	0.000
I2	0.649	0.048	13.608	0.000
I3	0.671	0.049	13.739	0.000
I4	2.445	0.133	18.389	0.000
I5	2.038	0.107	18.984	0.000
I6	1.071	0.111	9.639	0.000
I7	1.594	0.076	21.104	0.000
I8	0.609	0.078	7.815	0.000
I9	1.104	0.081	13.701	0.000
I10	0.725	0.044	16.658	0.000
I11	0.327	0.030	10.867	0.000
I12	0.713	0.044	16.204	0.000

The R^2 values below are presented at the end of the standardized solution and measure the proportion of variance in each observed variable that is explained by the latent variables. High values indicate that the observed variables are strongly related to the latent variables. From the values below we can see that I5 and I7 have low R^2 values, indicating they are not strongly related to their respective factors.

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
I1	0.659	0.023	29.228	0.000
I2	0.722	0.022	32.830	0.000
I3	0.722	0.022	33.094	0.000
I4	0.349	0.032	11.015	0.000
I5	0.277	0.032	8.786	0.000
I6	0.636	0.039	16.429	0.000
I7	0.213	0.026	8.082	0.000
I8	0.731	0.035	20.637	0.000
I9	0.571	0.032	17.787	0.000
I10	0.478	0.031	15.417	0.000
I11	0.658	0.033	20.167	0.000
I12	0.502	0.031	16.295	0.000

The standardized residuals are shown below. These are in a *z*-score metric, so values greater than |2.0| are typically taken as indications of misfit. Selected values are shown in Table 13.6 in the text, along with a discussion of possible reasons for the model misfit.

As can be seen from the table, many of the residuals are greater than |2.0|. This supports our earlier judgment of model misfit based on the fit indices.

Standardized Residuals (z-scores) for Covariances

	I1	I2	I3	I4	I5
I1	0.000				
I2	0.004	0.000			
I3	4.252	-22.306	0.000		
I4	-3.995	-0.840	-3.669	0.000	
I5	-2.717	0.671	-2.615	4.495	0.000
I6	-3.574	5.594	5.079	3.018	-5.125
I7	-4.722	-0.951	-2.709	3.291	7.180
I8	-1.384	2.611	2.936	-1.860	2.656
I9	-3.296	-0.409	0.528	-5.034	2.424
I10	4.900	3.666	2.219	0.151	1.270
I11	-0.920	-0.324	-1.308	-1.214	-1.717
I12	-2.898	-1.355	-2.667	-0.277	-0.417

Standardized Residuals (z-scores) for Covariances

	I6	I7	I8	I9	I10
I6	999.000				
I7	2.189	0.000			
I8	-1.638	-1.224	0.000		
I9	-0.090	-1.268	999.000	0.000	
I10	0.566	-2.759	-1.173	-1.180	0.000
I11	-0.266	-2.241	-0.284	-0.245	-0.711
I12	1.696	2.244	2.114	2.076	-1.025

```
    Standardized Residuals (z-scores) for Covariances
               I11           I12

               ________      ________
 I11             0.000
 I12             4.701         0.000
```

The model modification indices are presented last and are shown below. By default, M*plus* prints values greater than 10.0. To obtain all the modification indices, use the syntax **modindices (all)***.*

The first column refers to the parameter to be added to the model. In this column ""BY Statements" refer to loadings. For example, "PERFAPP BY I4" refers to the addition of the loading of I4 on PERFAPP. "WITH Statements" refer to covariances between pairs of residual variances. For example, "I3 WITH I1" refers to the covariance between the residuals of I3 and I1.

The second column, labeled "M.I." refers to the value by which the chi-square value would decrease if the parameter in column 1 were to be added to the model (the modification index).

In the third column, the "E.P.C." or Expected Parameter Change, is the estimated value of the parameter in question, if that parameter were to be added to the model.

The "Std E.P.C." and the StdYX E.P.C." are standardized versions of the EPC. In the "Std" version, the EPC is standardized based on the variances of the latent variables only, whereas in the "StdYX" version the EPC is standardized using the variances of both the latent and observed variables.

For example, the M.I. of 12.911 for "PERFAPP BY I4" indicates that, if I4 were allowed to load on PERFAPP the overall chi-square value would decrease by 12.911. The estimated value of the new loading would be -.229 in the unstandardized solution, -.287 in the Std standardized solution, and -.148 in the completely standardized (StdYX) solution.

MODEL MODIFICATION INDICES

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

	M.I.	E.P.C.	Std E.P.C.	StdYX E.P.C.
BY Statements				
PERFAPP BY I4	12.911	-0.229	-0.287	-0.148
PERFAPP BY I6	22.757	0.326	0.409	0.239
PERFAPP BY I10	20.215	0.126	0.158	0.135
PERAVOID BY I1	30.608	-0.235	-0.269	-0.174
PERAVOID BY I2	15.397	0.161	0.185	0.121
PERAVOID BY I7	13.124	0.185	0.212	0.149
MASAVOID BY I1	16.106	-0.234	-0.154	-0.099
MASAVOID BY I5	15.097	0.406	0.267	0.159
WITH Statements				
I3 WITH I1	17.588	0.311	0.311	0.420
I3 WITH I2	19.874	-0.363	-0.363	-0.551
I4 WITH I3	10.563	-0.170	-0.170	-0.132
I5 WITH I4	17.395	0.378	0.378	0.169
I6 WITH I1	12.569	-0.157	-0.157	-0.168
I6 WITH I3	15.912	0.172	0.172	0.203
I6 WITH I5	33.827	-0.591	-0.591	-0.400
I7 WITH I1	13.436	-0.153	-0.153	-0.134
I7 WITH I4	11.602	0.231	0.231	0.117
I7 WITH I5	40.375	0.386	0.386	0.214
I9 WITH I4	23.206	-0.310	-0.310	-0.189
I9 WITH I8	12.798	0.650	0.650	0.793
I10 WITH I1	19.295	0.133	0.133	0.174
I12 WITH I7	16.777	0.157	0.157	0.147