

# Quick Guide for Using Mplus

ESSENTIALS FOR GETTING STARTED WITH MPLUS

PC USERS

N. K. BOWEN, 2015

# Quick Guide for Using Mplus

*Disclaimer: Mplus is a powerful SEM program. Many options available in Mplus are not covered here. Information on the options that are covered is based on our experiences with recent versions of the program. These guidelines are not meant to be comprehensive or exhaustive. They reflect strategies that have worked for us. They may not reflect upcoming versions of the program or recommendations of the Mplus program developers.*

See also the Mplus website: <http://www.statmodel.com/index.shtml> for the *Mplus User's Guide* program updates, discussion board, and other resources, and consult the in-program help feature.

# Table of Contents

<b>1 PREPARING AND SAVING DATA FILES FOR USE WITH MPLUS</b>	<b>3</b>
1a Saving Data Files for Use in Mplus	3
1b More on Missing Values	4
<b>2 SPECIFYING DATA FILES AND VARIABLES IN MPLUS SYNTAX WITH THE DATA AND VARIABLE COMMANDS AND OPTIONS</b>	<b>5</b>
2a Specifying Data Files	5
2b Specifying Variables	6
<b>3 SPECIFYING ANALYSIS OPTIONS</b>	<b>8</b>
<b>4 MODEL SPECIFICATION</b>	<b>8</b>
13 Examples of Mplus Syntax for Measurement and General Structural Models	9
Example 4.1 3-factor CFA with 9 continuous, normally distributed observed variables, no missing values	9
Example 4.2 3-factor CFA with 9 continuous, normally distributed observed variables, and <u>missing values</u>	11
Example 4.3 3-factor CFA with 9 continuous, normally distributed observed variables, missing values, and <u>two correlated measurement errors</u>	12
Example 4.4 3-factor CFA with 9 continuous, <u>non-normal observed variables</u> , and missing values	13
Example 4.5 3-factor CFA with 9 <u>categorical observed variables</u> , and missing values	13
Example 4.6 CFA with 9 categorical observed variables, missing values, and <u>clustered data</u>	14
Example 4.7 CFA with Categorical and Clustered Data, and <u>Code for Chi Square Difference Test—1</u>	15
Example 4.8 CFA with Categorical and Clustered Data, and <u>Code for Chi Square Difference Test—2</u>	16
Example 4.9 Second Order Factor Model	17
Example 4.10 Multiple Group CFA	18
Example 4.11 General SEM Specification	19
Example 4.12 General SEM with Latent and <u>Observed Predictors</u>	21
Example 4.13 General SEM with Latent and Observed Predictors and <u>a Mediational Hypothesis</u>	22
<b>5 MPLUS' LANGUAGE GENERATOR</b>	<b>23</b>

# 1 PREPARING AND SAVING DATA FILES FOR USE WITH MPLUS

This section provides information on preparing data files for Mplus. Mplus has many options related to data and variables. The presentation is not exhaustive; we cover just some of the most common preparations social work researchers may need to make in order to have their files read properly by Mplus. Mplus reads text files.

Note that Mplus will save output in an output file with the same name as an input file. If you change a model and want to save a new output file, save the changed input file under a new name or your original output will be over written.

## 1a Saving Data Files for Use in Mplus

Before saving an input data file, all data cleaning should be completed, any necessary data transformations and recodes should be completed, and missing values should be recoded (see next section). Files may contain variables that will not be included in the currently planned Mplus analyses, but all variables in the text file will have to be named and listed in the Mplus syntax in order for the file to be read correctly by Mplus (more information is provided below).

Mplus only reads the first 8 letters in variables names. To avoid getting a warning that some variable names are too long, be sure that variable names listed in Mplus syntax have 8 letters or fewer.

Raw data files with only numeric variables should be saved as free or fixed ASCII files with extensions as part of their names. They may be saved them as tab, space, or comma delimited text files. If files are saved in free format, it is not necessary to specify the file format in Mplus syntax. Those saved in fixed format need to be described on the FORMAT line in Fortran notation (e.g., F8.2). There can be no blanks in files in free format (therefore, missing values cannot be designated with blanks in free format text files). Files saved from SPSS as tab delimited .dat files are among those read in Mplus. From Excel, files saved as text MS\_DOS files (with .txt extension), or “formatted text, space delimited” files (with .prn extension) are among those that can be read.

To use a lower triangle or full covariance matrix as the input matrix for Mplus, create a free format text file from a spreadsheet or database program with the matrix beginning on the first row of the file. To use a correlation matrix, create a free format text file that includes the means of the variables (in the same order as they occur in the correlation matrix) on the first line of the file, the standard deviations of the variables on the second line, and the first line of the full or lower triangle correlation matrix on the third line.

Example:

1. How to format a correlation matrix of 4 variables to be used in Mplus analyses.

.46	.34	.75	.52	←	Means of the four variables**
.14	.22	.49	.38	←	Standard deviations of the four variables
1					
.13	1			←	Correlation matrix of the four variables
.23	.81	1			
.48	.56	.44	1		

It is possible to create new variables from existing variables using simply arithmetic operations once you are in Mplus, but creating recoded variables before saving your text files saves you time later when you would have to enter the recode information in the syntax file each time you want to use the recoded variable.

\*Therefore, we recommend completing all recodes before saving your data into a text file for use in Mplus.\*

\*We recommend using files with .txt or .dat extensions. Do not add the extensions manually in Windows explorer. Sometimes no extension is shown on file names listed in Windows explorer. When you add an extension manually, you are actually giving the file a double extension (e.g., filename.txt.txt). If you add the extension manually, Mplus will be looking for a file named filename.txt and will tell you it cannot find it because it only sees filename.txt.txt. Find a SAVE option that creates a file with a .dat or .txt extension or, if no extension is shown, that is described as a text or dat file.\*

\*To ensure that all variables in your dataset are included in your list in Mplus and are in the correct order, we recommend copying the full list of variables from your general statistics program into your buffer so it can be pasted into Mplus syntax later. For example, using SPSS's pulldowns, request frequencies on all variables in the file. Paste the syntax, copy the list of variables, and then paste it into Mplus.\*

## **1b More on Missing Values**

As part of the preparation of data for SEM analysis in Mplus, users must designate which symbols or numbers in their datasets represent missing values. Options for missing values include: period (.), asterisk (\*), blank (), and numeric values that are not among valid options for a variable. Because of restrictions on the use of non-numeric flags (e.g., only one per dataset), we recommend using numeric flags (e.g., 99). In addition, with data in free format, blanks and periods may not be read correctly. Either positive or negative numeric values can be used as missing value flags; just be sure that the value used for any one variable does not overlap with its potential valid values. In Mplus, more than one missing flag may apply to one variable, one missing value flag can be used for all variables, or different flags can be used to designate missing values in different variables. Therefore, users do not have to alter existing data files to make all missing values the same. Different treatments of variable response options can be specified in Mplus without making changes to the original data file.

Note: By default, Mplus uses a Full Information Maximum Likelihood (FIML) estimation approach to handling missing values (if raw data are available and variables are treated as interval level or continuous). A discussion of missing data management is beyond the scope of this guide, but FIML is currently a highly recommended approach (e.g., Enders, C. K., 2010, *Applied missing data analysis*. New York: Guilford Press). As with all missing data approaches, it assumes data are not “missing not at random.”

\*We recommend recoding all missing values as 99, if possible. Then, as discussed below, under the VARIABLES command in the Mplus syntax file, one simply adds “Missing are all (99).”

## 2 SPECIFYING DATA FILES AND VARIABLES IN MPLUS SYNTAX WITH THE DATA AND VARIABLE COMMANDS AND OPTIONS

### 2a Specifying Data Files

Users specify where to find data files and what kinds of data are in data files with the DATA command and options. Note, the indentations of options under commands are for ease of illustration; they are not required in Mplus. Mplus code lines can be continued from one line to the next if necessary and end with a semi colon.

```
DATA:
    FILE IS          ;
    TYPE IS          ;
```

If a data file is not located in the same directory as the syntax file, path information is specified in the DATA: FILE IS line. Put the full directory information inside quotation marks to avoid potential problems with spaces in directory names. See Example 1 below.

If no TYPE IS line is included in the syntax, Mplus assumes the input file is in the common data format with one row for each subject and one column for each variable. If another format is used, TYPE IS must be included in the syntax. Other TYPE IS options include COVARIANCE and CORRELATION for lower triangle matrices, FULLCOV and FULLCORR for full covariance or correlation matrices, and IMPUTATION followed by a list of file names of imputed dataset.

Examples:

1. The FILE IS line below indicates that the free format text file called “worksatis.dat” can be found on the D drive in the folder called “Mplus analyses.” No TYPE is specified, so it is assumed that the data file has rows for records (subjects) and columns for variables.

```
DATA:
    FILE IS “D:\Mplus analyses\worksatis.dat”;
```

2. The FILE IS line below indicates that the free format text file called “Mplustest.dat” is in the current directory. The text file contains a lower triangle correlation matrix, preceded by a line (row) of means and a line of standard deviations. The third line indicates the number of observations (subjects) represented in the correlation matrix.

```
DATA:
    FILE IS agencyinfo.dat;
    TYPE IS CORRELATION MEANS STDEVIATIONS;
    NOBSERVATIONS ARE 390;
```

## 2b Specifying Variables

Users indicate characteristics of their variables in the VARIABLE command section of an Mplus input file.

```
VARIABLE:  
  NAMES ARE                ;  
  USEVARIABLES ARE         ;  
  MISSING ARE              ;  
  CATEGORICAL ARE         ;
```

The first “option” under the VARIABLE command is the required option of specifying the names of the variables included in the input dataset. The NAMES ARE line lists in order all the names of the variables in the data file to be analyzed. Mplus reads text files without names, so the names entered here do not need to match any variable names used previously or displayed in other programs. Names should be 8 or fewer characters long and start with letters, but can contain numbers and underscores. The VARIABLE: NAMES ARE line is where you could paste a list of your variables from your general statistics program instead of typing them all into Mplus.

Examples:

1. The NAMES ARE line in the example below indicates that the data file contains five variables. The USEVARIABLES ARE line indicates that only four of the five variables will be used in the current analysis.

```
VARIABLE:  
—————→ NAMES ARE gender age sowoexp typagen jobsatis;  
—————→ USEVARIABLES ARE age sowoexp typagen jobsatis;
```

2. The code below specifies that all of the variables in the input data file will be included in the subsequent modeling information. It is possible to list additional variables after ALL on the USEVARIABLE line if new variables have been created within the Mplus program (and therefore and not in the input file).

```
VARIABLE:  
—————→ NAMES ARE gender age sowoexp typagen jobsatis;  
—————→ USEVARIABLES ARE ALL;
```

The NAMES ARE and USEVARIABLES lines are critical! If the user attempts to model a variable later that was not included in the USEVARIABLES line, even if it appears in the NAMES ARE line, Mplus will return an error message. If a variable is listed in the USEVARIABLES line and not included in the later modeling statement, Mplus will inform the user that a variable it expected to use is “uncorrelated with all other variables.”

Missing values are described in the VARIABLE command based on how they were treated before the input data file was saved from another program. The examples below illustrate how one numeric missing value flag can be applied to all variables in a data file (Example 1),

different variables can be assigned different missing value flags (Example 2), and each variable can have more than one missing value flag (Example 2). Note that missing value information only applies to raw data files, not summary data like correlations.

Examples:

1. The MISSING line here indicates that for all variables, 99's should be read as missing values.

```
VARIABLE:  
  NAMES ARE gender age sowoexp typagen jobsatis;  
  USEVARIABLES ARE age sowoexp typagen jobsatis;  
—————> MISSING ARE ALL (99);
```

2. The MISSING line here indicates that missing values for variable 1 (v1) are indicated by 99, and that missing values for variable 2 (v2) are indicated by 0, 00, and 000.

```
VARIABLE:  
  NAMES ARE gender age sowoexp typagen jobsatis;  
  USEVARIABLES ARE ALL;  
—————> MISSING ARE v1 (99) v2 (0 00 000);
```

One of Mplus' strengths is its ability to appropriately analyze variables with different distributional and measurement qualities. The default assumption in the program is that variables are continuous. Users can specify non-continuous variables as CENSORED, CATEGORICAL, NOMINAL, OR COUNT variables if appropriate. We focus on the second and third of these types because they have been referenced throughout the book. The CATEGORICAL specification is for variables with between 2 and 10 ordered response options. CATEGORICAL is the appropriate designation for observed latent variable indicators that are measured with Likert scale or other ordinal measures. The NOMINAL designation is used for variables with between 2 and 10 unordered options. Variables that are specified as either CATEGORICAL or NOMINAL are recoded by Mplus such that the response option with the lowest level becomes 0, with other options assigned higher values in increment of 1.

Example:

1. In the example below, the CATEGORICAL and NOMINAL lines indicate that variables ordv1 through ordv5 are what we have referred to throughout this book as ordinal variables; gender is nominal. Age is being specified by default as a continuous variable.

```
VARIABLE:  
  NAMES ARE gender age ordv1 ordv2 ordv3 ordv4 ordv5;  
  USEVARIABLES ARE gender age ordv1 ordv2 ordv3 ordv4 ordv5;  
  MISSING ARE ALL (.);  
—————> CATEGORICAL ARE ordv1 ordv2 ordv3 ordv4 ordv5;  
—————> NOMINAL ARE gender;
```



### 3 SPECIFYING ANALYSIS OPTIONS

There are many analysis options in Mplus. They follow the ANALYSIS: command. For example:

```
ANALYSIS:  
  TYPE IS GENERAL;  
  ESTIMATOR IS ML;  
  ITERATIONS = 1000;  
  CONVERGENCE = .00005;
```

For many path models, CFA's and general SEM's the default TYPE IS GENERAL is adequate. As shown in the examples in the next section, however, other analysis types need to be specified for multilevel data. The default treatment of missing values in Mplus is now Full Information Maximum Likelihood. The default estimator is Maximum Likelihood. Similarly, there are default iteration (1000) and convergence criteria (.0005 or .00001). With a single level path, CFA, or general SEM, using variables that are interval level and relatively normally distributed, all of the defaults can be used and the ANALYSIS section can be left blank.

In many cases, social scientists have data that are not normally distributed or are ordinal. The examples in the next section illustrate how to request other estimators when data are not continuous and/or normally distributed.

The Iterations value indicates how many times the program will try out sets of parameter estimates in its attempt to minimize differences between the input matrix and the matrix implied by estimates. It is rarely necessary or useful to change the value. Most models converge with far fewer iterations, and if they don't, more iterations are not likely to solve the problem. The convergence criterion is the increment of improvement in the minimization function value at which the program stops seeking a better model. The program is said to have converged on a solution when tweaking parameters further leads to virtually no improvement in fit.

### 4 MODEL SPECIFICATION

Model specification occurs under the MODEL command. Here we provide annotated examples of a number of common types of CFA and general SEM models. Variations in the formatting of the code are presented to illustrate their equivalence. For example, "ARE," "IS" and "=" are equivalent. Commands can be spelled out in full or, in most cases, just the first four letters can be provided. Annotations provide more information on syntax and options.

## 13 EXAMPLES OF MPLUS SYNTAX FOR MEASUREMENT AND GENERAL STRUCTURAL MODELS

### Example 4.1 3-factor CFA with 9 continuous, normally distributed observed variables, no missing values

Measurement Model A. This is the CFA model specified in the measurement model examples below unless otherwise indicated:

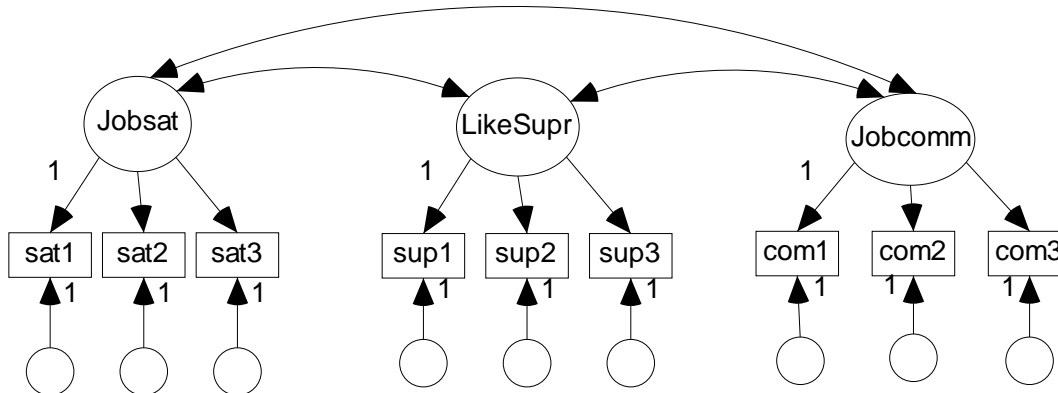


Figure 1. Measurement Model A

Mplus Code	Annotation
<p>TITLE: Measurement Model of Job Satisfaction and Commitment, Continuous, Normal Variables, No Missing Values</p> <p>DATA: FILE IS "C:\Mplus\jobdata.dat";</p> <p>VARIABLE: NAMES ARE agency sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3 age gender;</p> <p>USEVARIABLES ARE sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3;</p>	<p>Optional title is given; it will appear in output, which is useful</p> <p>Mplus reads text files; the default is free format, so it is not specified here.</p> <p>The dataset contains 12 variables.</p> <p>Only 9 of the variables will be used in this analysis.</p>
<p>ANALYSIS: ESTIMATOR IS ML</p> <p>ITERATIONS = 1000; CONVERGENCE = 0.00001;</p>	<p>ML is the default estimator; it appropriate for interval or higher normally-distributed variables</p> <p>These defaults numbers for iterations and convergence criterion can be increased if the model has trouble converging.</p>
<p>MODEL: JOBSATIS BY sat1 sat2 sat3 ; LIKESUPR BY sup1 - sup3; JOBCOMM BY com1 - com3;</p>	<p>The model is a 3-factor model with 3 observed indicators loading on each factor. "BY" indicates that JOBSATIS is being measured by SAT1-SAT3. Latent</p>

	factors are named here for the first time. The observed variable names must match the NAMES ARE and USEVARIABLE lines above. Correlations among the three factors are assumed by Mplus (default).
OUTPUT: STDYX MODINDICES (.5) RESIDUAL;	Fully standardized parameter estimates, modifications indices that are .5 or higher and the residual matrix are requested. Fit indices are provided by default.

**Note:** By default, the loading of the first indicator of a factor (first variable listed after BY) is fixed at 1; covariances between pairs of exogenous latent variables are freely estimated, and error paths are fixed at 1.

To change the default setting of the first loading of JOBSATIS to 1, you make a notation in the MODEL command to free the first loading and a notation to fix an alternative loading at 1 (here the loading of sat2 is fixed at 1). For example:

```
JOBSATIS BY sat1* sat2@1 sat3;
```

The “\*” indicates that the default fixing of sat1’s value to 1 is changed, and the “@ 1” specifies that the loading of sat2 is fixed at 1 instead. These specifications occur in the  $\Lambda\delta$  matrix.

To set the metric of JOBSATIS by fixing its variance to 1 instead of one of its loadings, similar code is used:

```
JOBSATIS BY sat1* sat2 sat3;  
JOBSATIS @ 1;
```

To change the default setting of an inter-factor correlation to 0, a “WITH” statement is added to the model. For example:

```
JOBSATIS WITH JOBCOMM @ 0;
```

This code specifies that the correlation between the two factors is 0. Inter-factor correlations are found on the off-diagonals in the  $\Phi$  matrix.

\*We recommend always examining output carefully to determine if it indicates an unexpected default setting was used. Output reflects the actual parameters that were and were not estimated. Be sure they are not different than what you intended.\*

\*We recommend using the STDYX option for standardized output. It returns fully standardized coefficients. With the STANDARDIZED option, three sets of standardized output are returned in the output.\*

**Example 4.2 3-factor CFA with 9 continuous, normally distributed observed variables, and missing values**

<b>Mplus Code</b>	<b>Annotation</b>
TITLE: Measurement Model of Job Satisfaction and Commitment with Continuous, Normal Variables, and Missing Values DATA: FILE IS "C:\Mplus\jobdata.dat";	
VARIABLE: NAMES ARE agency sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3 age gender; USEVARIABLES ARE sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3; MISSING ARE ALL (99);	The missing value flag for all variables in the current analysis is 99.
ANALYSIS: TYPE IS MISSING; ESTIMATOR IS ML; ITERATIONS = 1000; CONVERGENCE = 0.00001;	The new TYPE line request that missing values be handled by FIML. FIML is now used by default so this line does not need to be in the syntax. (But note that FIML cannot be used with summary data or categorical variables.)
MODEL: <i>Same as previous</i>	
OUTPUT: SAMPSTATS STDYX MODINDICES (1) RESIDUAL;	Statistics on the input variables, fully standardized parameters, modifications indices that change $\chi^2$ by more than 1 are requested, and the residual matrix are requested as output.

**Example 4.3 3-factor CFA with 9 continuous, normally distributed observed variables, missing values, and two correlated measurement errors**

Measurement Model B. The CFA model specified in Example #3 has two correlated measurement errors as pictured here:

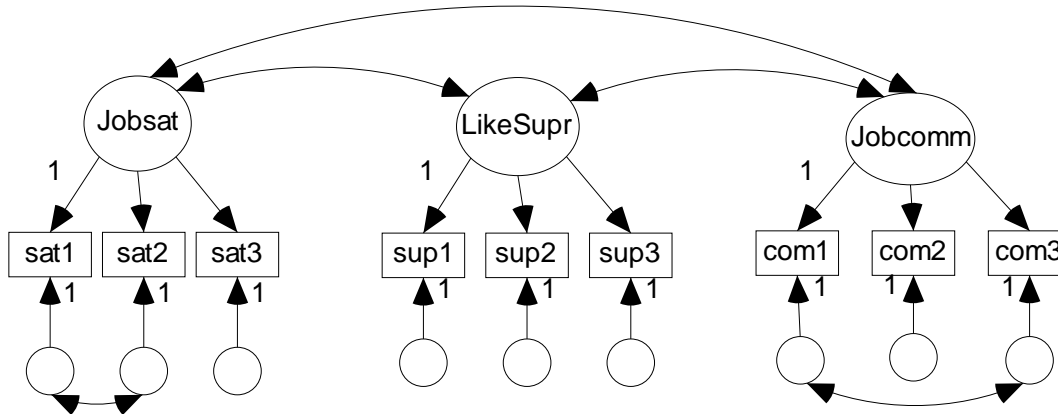


Figure 2. Measurement Model B

The code for this example is the same as the code for the previous example except for the addition of two lines to the MODEL statement. Only part of the code is presented.

Mplus Code	Annotation
TITLE: Measurement Model of Job Satisfaction and Commitment with Missing Values, Continuous Variables, Two Correlated Errors DATA: FILE IS "C:\Mplus\jobdata.dat";	
MODEL: JOBSATIS BY sat1 sat2 sat3 ; LIKESUPR BY sup1 - sup3; JOBCOMM BY com1 - com3; sat1 WITH sat2; com1 WITH com3;	Each factor has three observed indicators. One pair of indicators of JOBSATIS and one pair of indicators for JOBCOMM have correlated measurement errors.

The two “WITH” statements (in conjunction with the BY statements that specify the correlated variables as indicators of latent variables) specify that the two pairs of measurement error variances are correlated (not the variables themselves). The estimated parameters will be in the  $\Theta_6$  matrix.

**Example 4.4 3-factor CFA with 9 continuous, non-normal observed variables, and missing values**

The code for this example would be the same as the code for the models above except for a change in the ESTIMATOR option.

<b>Mplus Code</b>	<b>Annotation</b>
TITLE: Measurement Model of Job Satisfaction and Commitment with Non Normal Continuous Variables and Missing Values DATA: FILE IS "C:\Mplus\jobdata.dat";	
ANALYSIS: ESTIMATOR IS MLM; ITERATIONS = 1000; CONVERGENCE = 0.00001;	The MLM estimator corrects standard errors and the chi square statistic for non-normality.

**Example 4.5 3-factor CFA with 9 categorical observed variables, and missing values**

The code for this example is the same as the code for the models above except for changes in the VARIABLE Command and ESTIMATOR option.

<b>Mplus Code</b>	<b>Annotation</b>
TITLE: Measurement Model of Job Satisfaction and Commitment, Categorical Variables and Missing Values, DATA: FILE IS "C:\Mplus\jobdata.dat";	
VARIABLE: NAMES ARE agency sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3 age gender; USEVARIABLES ARE sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3; MISSING ARE ALL (99); CATEGORICAL ARE sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3;	This line indicates that the observed indicators of the latent variable are categorical (ordinal) variables.
ANALYSIS:  ESTIMATOR IS WLSMV; ITERATIONS = 1000; CONVERGENCE = 0.00001;	The WLSMV estimator creates and analyses a matrix of polychoric, tetrachoric, and/or polyserial correlations and an associated weight matrix. When variables are labeled as categorical, Mplus uses WLSMV by default. You cannot use FIML with summary data, so Mplus uses a pairwise deletion approach. See

	Asparouhov & Muthen, 2010, “Weighted least squares estimation with missing data,” at the Mplus website for more information).
MODEL: JOBSATIS BY sat1 sat2 sat3 ; LIKESUPR BY sup1 – sup3; JOBCOMM BY com1 - com3;	
OUTPUT: SAMPSTATS STDYX MODINDICES (1) RESIDUAL;	A smaller number of fit indices is available with WLSMV. They are provided by default in the output.

**Note:** This code illustrates one proper way to estimate a CFA with ordinal variables. If categorical variables are specified and the ML or MLM estimator is chosen, Mplus will change the estimator to WLSMV (and let you know with a warning in the output).

**Example 4.6 CFA with 9 categorical observed variables, missing values, and clustered data**

TITLE: Measurement Model of Job Satisfaction and Commitment, Categorical Variables, Missing Values, and Clustered Data DATA: FILE IS "C:\Mplus\jobdata.dat";	
VARIABLE: NAMES ARE agency sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3 age gender; USEVARIABLES ARE sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3; MISSING ARE ALL (99); CATEGORICAL ARE sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3; CLUSTER IS agency;	Don't include agency on the USEVARIABLES line.  Data on individuals are clustered within the variable “agency.” Agency is not included in the MODEL command; its role is specified here.
ANALYSIS: TYPE IS COMPLEX; ESTIMATOR IS WLSMV; ITERATIONS = 1000; CONVERGENCE = 0.00001;	With TYPE =COMPLEX, standard errors and chi square will be corrected for clustering. The categorical nature of the data is handled by WLSMV.
MODEL: JOBSATIS BY sat1 sat2 sat3 ; LIKESUPR BY sup1 – sup3; JOBCOMM BY com1 - com3;	
OUTPUT: STDYX MODINDICES (.5) RESIDUAL;	

**Example 4.7 CFA with Categorical and Clustered Data, and Code for Chi Square Difference Test—Step 1 (See Example 4.8 for Step 2)**

When users use certain estimators in Mplus, they will see the following notice under the  $\chi^2$  part of the output: “The chi-square value for MLM, MLMV, MLR, ULS, WLSM and WLSMV cannot be used for chi-square difference tests.” Both the  $\chi^2$  and the *df* are calculated differently with these estimators. Users will note, for example, that the *df* is not the difference between the number of unique sample moments (input covariance matrix elements) and the number of parameters being estimated. To compare nested models when the above estimators are used in Mplus requires a special process. The user cannot just compare the change in  $\chi^2$  per change in *df*. First the user runs the less restrictive model, and includes the SAVEDATA command.

TITLE: Measurement Model of Job Satisfaction and Commitment, Categorical Variables, Missing Values, Clustered Data, and Difftest SAVEDATA option DATA: FILE IS “C:\Mplus\jobdata.dat”;	
VARIABLE: <i>Same as previous</i>	
ANALYSIS: TYPE IS COMPLEX; ESTIMATOR IS WLSMV; ITERATIONS = 1000; CONVERGENCE = 0.00001;	With TYPE =COMPLEX, standard errors and chi square will be corrected for clustering. The categorical nature of the data is handled by WLSMV.
MODEL: <i>Same as previous</i>	
OUTPUT: STDYX MODINDICES (.5) RESIDUAL;	
SAVEDATA: DIFFTEST IS jobsat1;	A small text file will be saved in the input file directory for use by the program in computing the difference test when the nested model is run.



**Example 4.8 CFA with Categorical and Clustered Data, and Code for Chi Square Difference Test—Step 2 (See Example 4.7 for Step 1)**

Next the user runs the nested model (with fewer parameters freely estimated than in the previous model). A new line is added to the ANALYSIS options.

TITLE: Measurement Model of Job Satisfaction and Commitment, Categorical Variables, Missing Values, Clustered Data, and DIFFTEST Analysis Option DATA: FILE IS "C:\Mplus\jobdata.dat";	
VARIABLE: <i>Same as previous</i>	
ANALYSIS: TYPE IS COMPLEX; ESTIMATOR IS WLSMV; ITERATIONS = 1000; CONVERGENCE = 0.00001; DIFFTEST=jobsat1;	The last option under the ANALYSIS command tells the program to use the data saved in file created when the first model was run, to test whether the fit of the more restrictive (second) model is significantly worse than the fit of the first.
MODEL: <i>Any model nested in previous model</i>	
OUTPUT: STDYX MODINDICES (.5) RESIDUAL;	

The output for the second model will include results of the difference test. If the  $\chi^2$  difference has a  $p$  value less than .05, the restrictions added to create the nested model make the fit significantly worse and the first model is retained. If  $\chi^2$  did not become significantly worse ( $p > .05$ ), the second, more parsimonious, model should be retained.

### Example 4.9 Second Order Factor Model

Measurement Model C. The diagram here shows a higher order factor “Work Experience” that influences the three lower order factors:

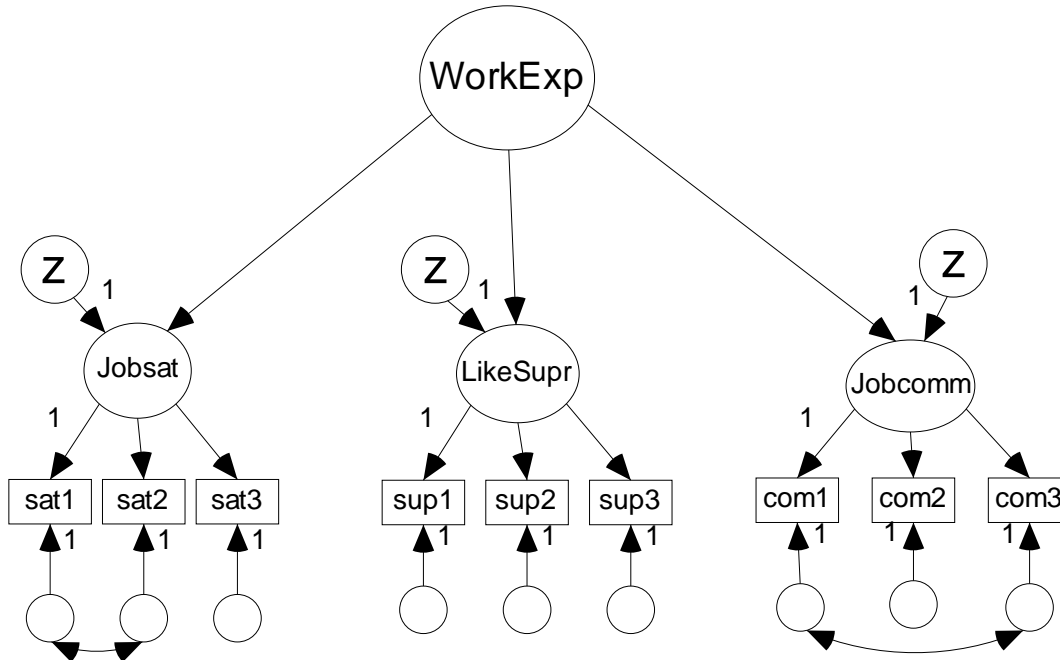


Figure 3. Measurement Model C

TITLE: Second Order Factor Model with Categorical Variables and Missing Values	
DATA: FILE IS "C:\Mplus\jobdata.dat";	
VARIABLE: NAMES ARE agency sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3 age gender; USEVARIABLES ARE sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3; MISSING ARE ALL (99); CATEGORICAL ARE sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3;	
ANALYSIS: ESTIMATOR IS WLSMV; ITERATIONS = 1000; CONVERGENCE = 0.00001;	
MODEL: JOBSATIS BY sat1 sat2 sat3 ; LIKESUPR BY sup1 - sup3; JOBCOMM BY com1 - com3;	The new line of code names a new latent factor and specifies that it is measured by the first three latent

WORKEXP BY JOBSATIS LIKESUPR JOBCOM; Sat1 with sat2; Com1 with com3;	factors. The first order factors become endogenous variables.
OUTPUT: STDYX MODINDICES (.5) RESIDUAL;	

#### Example 4.10 Multiple Group CFA

A measurement model (or general SEM) can be tested to see if it is “invariant” across groups. Additional information is required in the VARIABLE, ANALYSIS, and MODEL commands. Multiple group model instructions vary depending on a number of variable and model issues. The example here is simple. A more detailed example is provided in the online book resources. Readers are also referred to the most recent online *Mplus User’s Guide* for more information.

TITLE: Multiple Group CFA with Categorical Variables and Missing Values DATA: FILE IS "C:\Mplus\jobdata.dat";	
VARIABLE: NAMES ARE agency sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3 age gender; USEVARIABLES ARE sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3; MISSING ARE ALL (99); CATEGORICAL ARE sat1 sat2 sat3 sup1 sup2 sup3 com1 com2 com3; GROUPING IS gender (0=women 1=men)	
ANALYSIS: MODEL = NOMEANSTRUCTURE; ESTIMATOR IS WLSMV; ITERATIONS = 1000; CONVERGENCE = 0.00001;	
MODEL: JOBSATIS BY sat1 sat2 sat3 ; LIKESUPR BY sup1 – sup3; JOBCOMM BY com1 - com3; Sat1 with sat2; Com1 with com3;  MODEL men: sat1 with sat2 @0; com1 with com3@0;	The model statement specifies the model that applies to both groups.  The MODEL men statement indicates where the model differs across the two groups. In this example, the correlated measurement errors do not apply to the men’s model.

### Example 4.11 General SEM Specification

Most aspects of model specification are the same for general SEM as for CFA. The same reasons for choosing different estimators and analysis options apply. Multiple group and multi-level modeling can be done. See the measurement model examples above for information on syntax that applies to both types of models. Differences will be in the MODEL command, where relationships between factors are specified for general SEMs. After the first general SEM example, just the TITLE and MODEL command sections are presented for different structural models.

General SEM A. The model pictured below is based on the measurement model used in the examples above. Note that the structural part of the model is just-identified ( $df = 0$ ). In a real study, we would have to address this issue in order to test the hypothesized relationships among latent variables. General SEM B shows one way we could make the model over-identified.

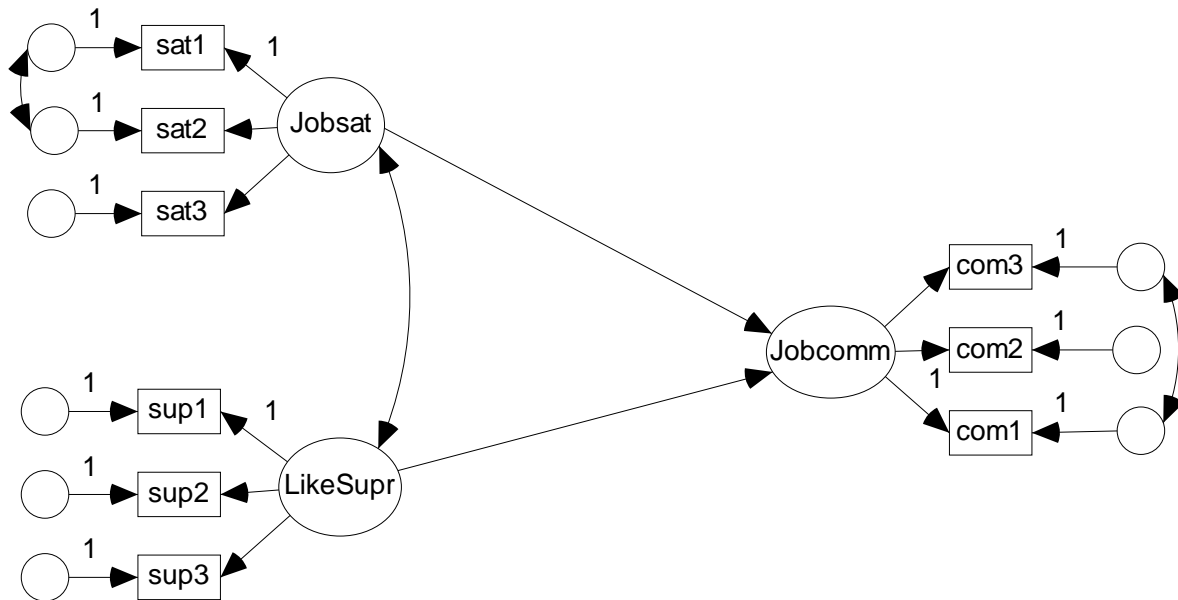


Figure 4. General SEM A

<p>TITLE: General SEM of Job Satisfaction and Commitment with Categorical Variables, Missing Values, Clustered Data          DATA: FILE IS "C:\Mplus\jobdata.dat";</p>	
<p>VARIABLE:          NAMES ARE              agency sat1 sat2 sat3 sup1 sup2              sup3 com1 com2 com3 age gender;          USEVARIABLES ARE sat1 sat2 sat3 sup1              sup2 sup3 com1 com2 com3;          MISSING ARE ALL (99);</p>	

<p>CATEGORICAL ARE sat1 sat2 sat3 sup1  sup2 sup3 com1 com2 com3;  CLUSTER IS agency;</p>	
<p>ANALYSIS:  TYPE IS COMPLEX;  ESTIMATOR IS WLSMV;  ITERATIONS = 1000;  CONVERGENCE = 0.00001</p>	
<p>MODEL:  JOBSATIS BY sat1 sat2 sat3 ;  LIKESUP BY sup1 – sup3;  JOBCOMM BY com1 - com3;  sat1 WITH sat2;  com1 WITH com3;  JOBCOMM ON JOBSATIS;  JOBCOMM ON LIKESUPR;</p>	<p>The measurement model is specified as before.  Here we have included the correlated errors of measurement presented earlier.  The ON statements indicate regression relationships among the latent variables.  The correlation between JOBSATIS and LIKESUPR is assumed by Mplus.</p>
<p>OUTPUT: STDYX MODINDICES (.5) RESIDUAL;</p>	

**Example 4.12 General SEM with Latent and Observed Predictors**

General SEM B. The inclusion of gender as a predictor of Job Commitment makes the structural part of the model over-identified ( $df = 2$ ).

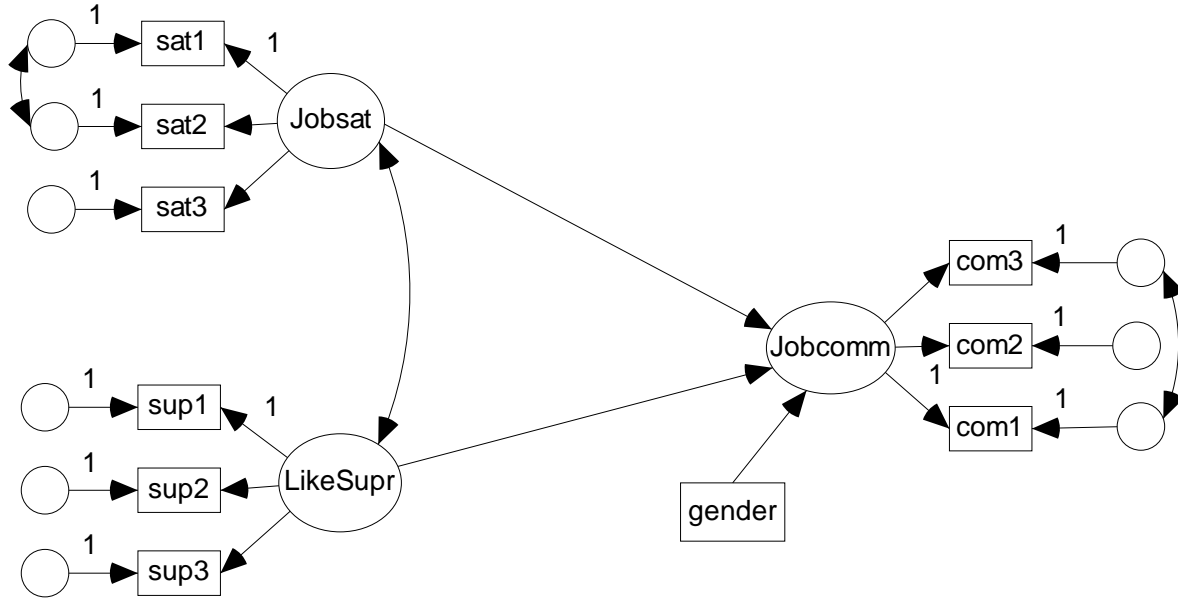


Figure 5. General SEM B

<p>TITLE: General SEM of Job Satisfaction and Commitment with Categorical Variables, Missing Values, Clustered Data          DATA: FILE IS "C:\Mplus\jobdata.dat";</p>	
<p>MODEL:          JOBSATIS BY sat1 sat2 sat3 ;          LIKESUP BY sup1 - sup3;          JOBCOMM BY com1 - com3;          sat1 WITH sat2;          com1 WITH com3;          JOBCOMM ON JOBSATIS;          JOBCOMM ON LIKESUP;          JOBCOMM ON gender;</p>	<p>The new ON statement indicates that JOBCOMM is also regressed on gender.</p>

**Example 4.13 General SEM with Latent and Observed Predictors and a Mediational Hypothesis**

General SEM C. In the following SEM, Jobsat has both direct and indirect effects on Jobcomm. Indirect effects are through LikeSupr.

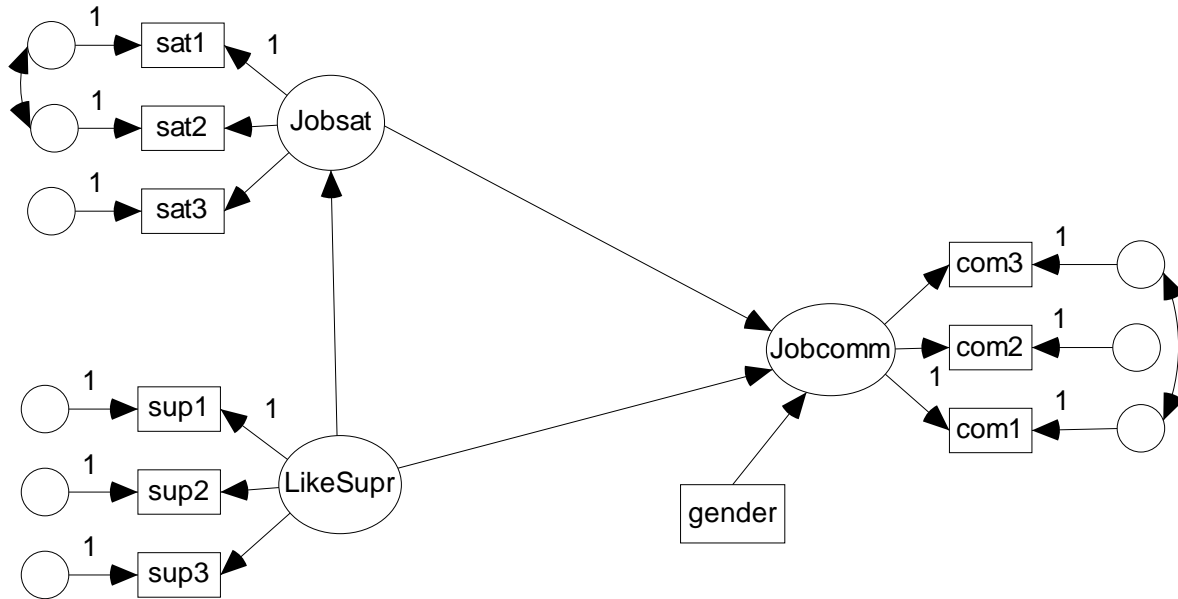


Figure 6. General SEM C

<p>TITLE: General SEM of Job Satisfaction and Commitment with Categorical Variables, Missing Values, Clustered Data, and Mediation          DATA: FILE IS "C:\Mplus\jobdata.dat";</p>	
<p>MODEL:          JOBSATIS BY sat1 sat2 sat3 ;          LIKESUP BY sup1 – sup3;          JOBCOMM BY com1 - com3;          sat1 WITH sat2;          com1 WITH com3;          JOBCOMM ON JOBSATIS;          JOBCOMM ON LIKESUP;          JOBCOMM ON gender;          JOBSATIS ON LIKESUPR;</p>	<p>The new ON statement indicates that JOBSATIS is regressed on LIKESUPR. JOBSATIS is both predicted and a predictor in the model. The implication is that the effects of LIKESUP on JOBCOMM are partially mediated by its effect on JOBSATIS.</p>

## 5 MPLUS' LANGUAGE GENERATOR

Mplus has a language generator that facilitates the specification of data, variables, and estimation options. The language generator allows the user to generate part of the code for an analysis quickly and accurately. The language generator does not create code specifying the model; that code is added later by the user to the generated syntax. The choices that appear throughout the language generator sequence vary slightly depending on your preliminary choice of SEM, SEM with missing data, and SEM with clustered data. For example, if you select SEM, you will see the option of indicating that you have “individual” (raw) data or “summary” data (a covariance or correlation matrix). This option does not appear if you choose SEM with missing or clustered data because summary data cannot be used for those types of analyses. Here we provide a quick tour of the language generator as it would apply to an analysis of clustered data with missing values. The text created from the language generated can be edited and amended after it is pasted. Some users may prefer to start their syntax files with the generator; others may prefer writing their own code from scratch.

Go to the Mplus pulldown menu at the top of the screen you see when Mplus opens. Choose “language generator.” Choose “SEM with clustered data.” (Even if you plan to do an SEM with clustered data AND missing data, select “SEM with clustered data.”)

Enter a title for the analysis you are about to do if you would like to have a title on your input code. For this example, we enter: Example of Language Created with Language Generator: SEM with Clustered Data. Browse for the input data file and select it. Click on Next. Indicate if the data file is in free or fixed format. (The examples we gave above assumed free format, and correlation and covariance matrix data must be in free format.) Indicate if there are missing values in the data, and whether they are identified with numeric or non-numeric values. Click Next.

In the next screen you are asked to enter variable names and add them to the list of variables in the data file one by one. If you have a dataset with many variables, we suggest entering just one or two variable names and cutting and pasting the rest into the code after the generated language is pasted into the Mplus editor. Click Next. In the next screen, the list of variables you have entered will appear on the left. Select the one variable (or more) you entered and move it into the USEVARIABLE box to the right. This is where the USEVARIABLE line is created. Add only those variables you plan to use in the current analysis. If you entered only one or two variable names at the previous screen, the rest can be pasted in later. Another option on this screen is to restrict the observations used to those with specific values on one or more variables. For example, if the proposed analysis is of individuals over 50 as indicated in a variable called “age,” specify in the box that “Age GE 50.” More than one restriction can be specified using AND, OR, NOT, and the following EQ (equal), NE (not equal), GE (greater than or equal to), LE (less than or equal to), LT (less than), GT (greater than). We do not specify and subset of observations for this example. We specify that we will be using the five ordinal variables in the analysis. We do not list “cluster” in the USEVARIABLE line; it will be given a special role as the clustering variable in the next screen.

In the next screen, you indicate what values represent missing values in your dataset. We specify that all variables use 99 as a missing variable flag. Click Next. In the subsequent screen, variables can be assigned special roles—the grouping variable in a multiple group analysis, the cluster variable in a multilevel model, variables measured at level 2 in a two level model (between variable), and weighting variable. Because this example assumes an analysis with



clustered data, we indicate that the variable called “cluster” in our dataset indicates how our level 1 subjects are grouped (e.g., within classrooms, agencies, states, etc.).

The following screen has analysis options. We choose “complex” because we want to correct for standard error biases created by the grouped nature of our data. (If you planned to model level 2 predictors, you would choose “twolevel.”) The default estimator for a complex analysis is MLM, which uses the Satorra Bentler chi square statistic corrected for clustering of cases. We leave the next two options related to optimization at their default levels. They can be altered if we have convergence problems later.

The final screen in the language generator asks what output we would like in addition to the default output. We indicate that we would like standardized parameter estimates. The code created by the Mplus language generator from the above example and pasted into the Mplus editor is as follows:

Example of Language Generator Syntax before Editing:

---

TITLE: Example of Language Created with Language Generator: SEM with Clustered Data

DATA:

FILE IS "F:\Mplus analyses\Langexample\Mydata.dat";

VARIABLE:

NAMES ARE age gender ordv1 ordv2 ordv3 ordv4 ordv5 cluster;

USEVARIABLES ARE ordv1 ordv2 ordv3 ordv4 ordv5;

MISSING ARE ALL (99);

CLUSTER IS cluster;

ANALYSIS:

TYPE IS COMPLEX;

ESTIMATOR IS MLM;

ITERATIONS = 1000;

CONVERGENCE = 0.00005;

OUTPUT: STANDARDIZED;

---

The code pasted into the Mplus editor by the language generator is now ready for editing and amendment. If there are more variables in the dataset than we entered in the generator, they can be added as appropriate to the NAMES ARE and USEVARIABLES ARE lines. The generator for “SEM with clustered data” did not give us the option of specifying categorical or nominal level variables. We add two lines under the VARIABLE command to specify the measurement level of our variables. The default estimator for SEM analyses with categorical variables is WLSMV, so we change MLM to WLSMV. (Mplus syntax will run when an inappropriate estimator is specified, but it will use the appropriate default estimator and notify the user of the change in the analysis output.) We also add a line under the ANALYSIS command specifying that we have missing values. FIML cannot be used with categorical data and WLSMV because that estimation approach uses summary data. MPlus uses a pairwise deletion approach with categorical data (see Asparouhov & Muthen, 2010, “Weighted least

squares estimation with missing data,” at the Mplus website for more information). Then we add a MODEL statement in between the ANALYSIS and OUTPUT commands. In this example we indicate that the five categorical variables load on one latent variable. Specification is discussed in more detail in the next section.

Example of Language Generator Syntax After Editing to Make it Complete:

---

TITLE: Example of Edited Syntax: SEM with Clustered Data

DATA:

FILE IS "F:\Mplus analyses\Langexample\Mydata.dat";

VARIABLE:

NAMES ARE age gender ordv1 ordv2 ordv3 ordv4 ordv5 cluster;

USEVARIABLES ARE ordv1 ordv2 ordv3 ordv4 ordv5;

MISSING ARE ALL (99);

CATEGORICAL ARE ordv1 ordv2 ordv3 ordv4 ordv5 ; ←5 variables are  
categorical.

NOMINAL IS gender; ← 1 variables is nominal.

ANALYSIS:

TYPE IS COMPLEX; ← We want correction of standard errors  
appropriate for our multilevel data

ESTIMATOR IS WLSMV; ← For categorical data; uses a  
polychoric correlation matrix

ITERATIONS = 1000; ← Tells Mplus to try up to 1000 times to  
achieve a converged solution (default)

CONVERGENCE = 0.00005; ← Defines convergence as one minimization  
value being this close to a prior iteration's  
minimization value

MODEL:

Factor1 by ordv1 ordv2 ordv3 ordv4 ordv5; ← Specifies the model to be tested:  
as our 5 observed variables loading  
on one new latent variable (Factor 1)

OUTPUT: STDYX;

← Requests fully standardized parameter estimates  
in addition to default output