

The analysis of multivariate data with missing values

Contents

1. Introduction.....	1
2. The Data	1
3. The Measurement Model.....	2
4. The Multiple Imputation Method for Data with Missing Values	3
5. The FIML Method for Data with Missing Values	7

1. Introduction

In practice, many multivariate data sets contain missing values. These missing values may result from non-responses in a survey, absenteeism of participants in a longitudinal study, etc. The traditional way of dealing with these missing data values is to use list wise deletion to generate a data set that only contains the complete data cases. However, list wise deletion may result in a very small data set. It is a well-known fact that most multivariate statistical methods require a large sample size, especially if the number of observed variables is large. Consequently, alternative statistical methods for dealing with data with missing values are of interest.

Multiple Imputation (MI) and Full Information Maximum Likelihood (FIML) estimation are two popular statistical methods for dealing with data with missing values. Both these methods are available in LISREL (Jöreskog & Sörbom 2003). The Multiple Imputation module of LISREL implements the Expected Maximization (EM) algorithm and the Markov Chain Monte Carlo (MCMC) method for imputing missing values in multivariate data sets. Technical details of these methods are available in Schafer (1997) and Du Toit & Du Toit (2001). Supplementary notes on these methods are also provided by Du Toit & Mels (2002).

In this note, the Multiple Imputation and FIML methods for data with missing values of LISREL are illustrated by fitting a measurement model to a multivariate data set consisting of the scores of a sample of girls on six psychological tests. This data set is described in the next section. The measurement model is described in section 3. Thereafter, the method of Multiple Imputation is used to fit the measurement model to the data set for girls. In section 5, the measurement model is fitted to the girls' data by means of the FIML method.

2. The Data

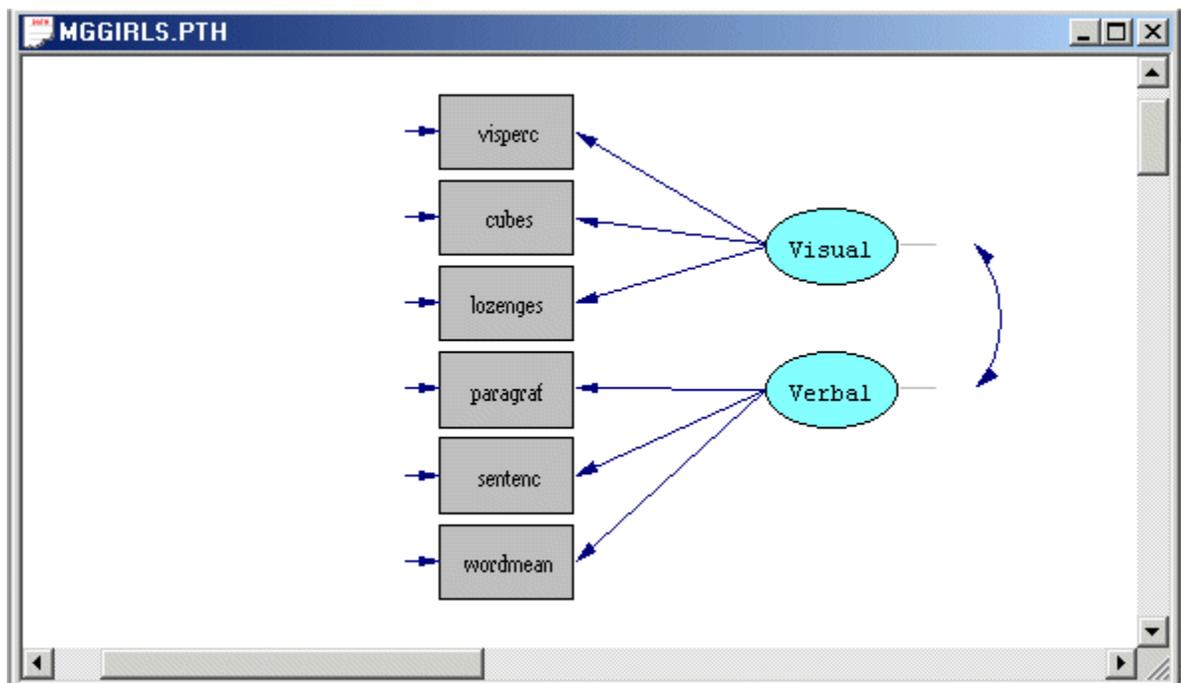
The data were simulated, with missing values at random, from a multivariate Normal distribution and are the scores of 1500 girls on six psychological tests. The raw data are listed in the LISREL System File (LSF) **MGGIRLS.LSF** in the **Missing Data Examples** subfolder of the LISREL installation folder. The first couple of lines of this file are shown in the following LSF window.

	visperc	cubes	lozenges	paragraf	sentenc	wordmean
1	-0.71	-1.96	-6.87	-9.00	-0.24	7.60
2	-6.42	-8.64	-4.37	0.59	1.53	9.63
3	8.79	-2.15	10.21	2.29	2.81	8.92
4	9.72	-2.38	14.83	0.48	4.17	6.80
5	-9.00	0.65	-8.02	-9.00	-7.51	-12.61
6	-10.57	-9.00	-2.18	0.48	6.70	-9.00
7	-9.00	7.80	-9.00	-9.00	-1.22	5.38
8	-9.00	-2.14	5.24	-1.01	2.72	3.73
9	2.60	5.72	-0.64	-0.04	-1.32	-9.00
10	0.73	-7.94	-6.89	-9.00	-9.00	-27.54
11	-3.66	-4.20	3.51	6.26	9.25	12.35
12	-6.26	0.88	-14.83	4.86	0.88	12.27
13	6.53	-9.00	-8.80	-9.00	-9.00	-1.79
14	8.77	-3.62	-9.00	-3.47	-0.57	-9.00
15	-2.61	-6.94	1.41	-0.65	-5.81	-9.00
16	2.14	-3.61	3.60	5.11	3.33	6.69
17	-8.59	-5.92	-10.00	3.70	2.84	6.02
18	-7.98	-0.13	-7.12	-4.13	-7.88	-7.12
19	11.09	-3.32	6.42	-2.06	-1.24	-5.31
20	2.19	0.49	-1.88	5.21	-9.00	7.14

Note that the -9.000 entries represent missing values.

3. The Measurement Model

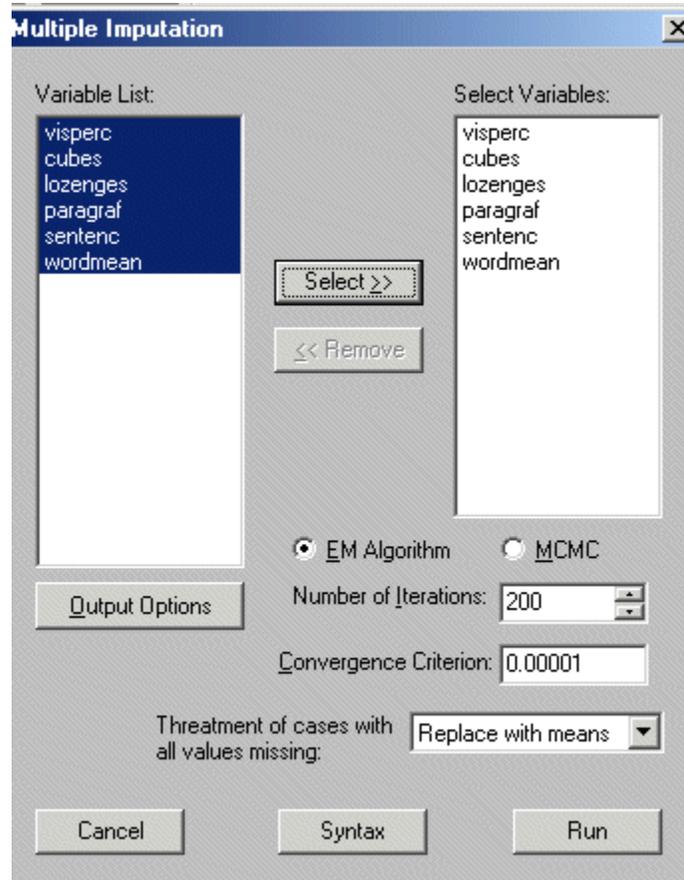
We consider data with missing values on six psychological tests. These six psychological tests are theoretically constructed to measure the visual perception and verbal ability of seventh- and eighth-grade children. A path diagram for the corresponding measurement model for verbal perception and verbal ability is shown in the figure below.



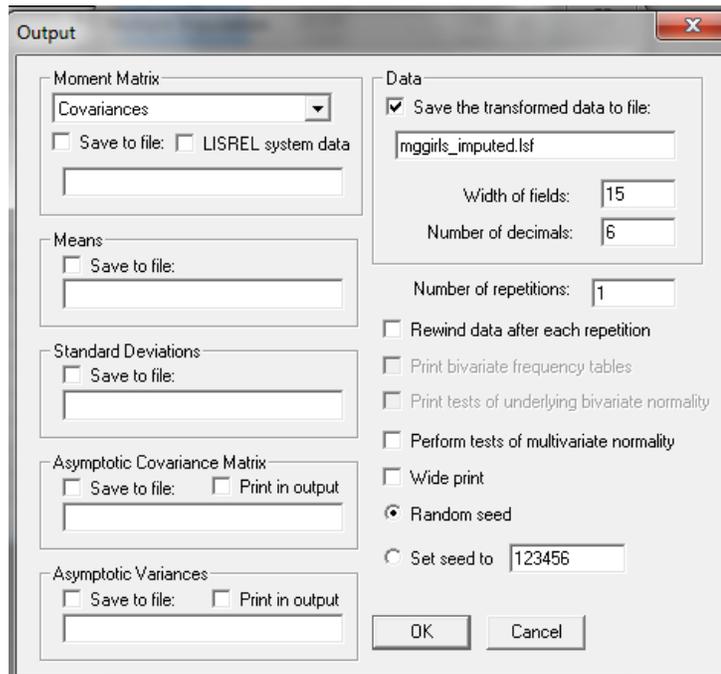
4. The Multiple Imputation Method for Data with Missing Values

Use the **Open** option on the **File** menu to load the **Open** dialog box. Next, browse for the **Missing Data Examples** folder of the LISREL installation folder. Select the **LISREL Data (*.lsf)** option from the **Files of type** dropdown list box and select the file **MGGIRLS.LSF**. Click on the **OK** push button to open the LSF window for **MGGIRLS.LSF**.

Select the **Multiple Imputation** option on the **Statistics** menu of the LSF window to load the **Multiple Imputation** dialog box. Select all the labels in the **Variable List** box. Click on the **Select** push button to produce the following **Multiple Imputation** dialog box.



Click on the **Output Options** push button to load the **Output** dialog box and check the **Save the transformed data to file** checkbox in the **Data** section. Enter the name **MGGIRLS_IMPUTED.LSF** in the string field of the **Data** section to produce the following **Output** dialog box.



Click **OK** on the **Output** dialog box to return to the **Multiple Imputation** dialog box. Click on the **Run** push button of the **Multiple Imputation** dialog box to produce the text editor window for **MGGIRLS.OUT**.

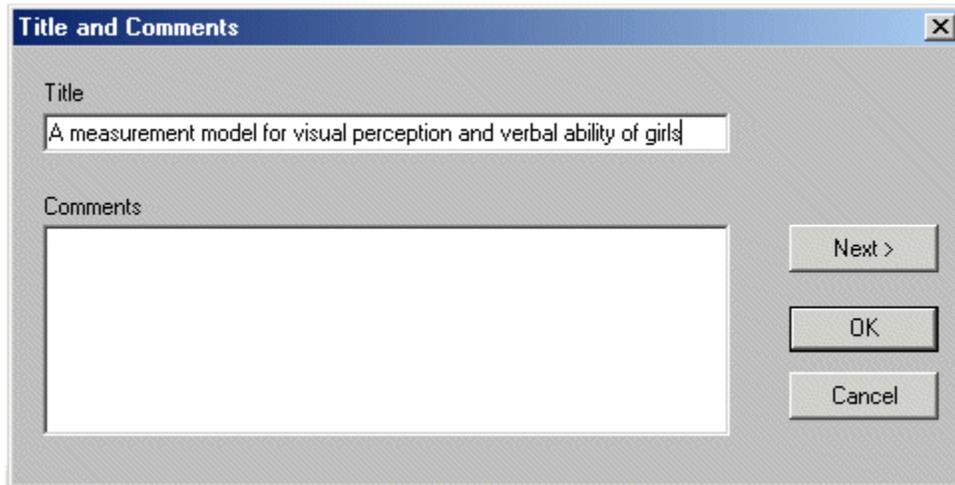
The imputed LSF **MGGIRLS_IMPULDED.LSF** is written to the **Missing Data Examples** folder. The first portion of this file is shown in the following LSF window.

	visperc	cubes	lozenges	paragraf	sentenc	wordmean	
1	-0.713	-1.957	-6.869	1.411	-0.238	7.601	
2	-6.419	-8.642	-4.371	0.591	1.530	9.626	
3	8.788	-2.154	10.208	2.290	2.809	8.916	
4	9.723	-2.379	14.827	0.476	4.165	6.803	
5	-2.813	0.649	-8.022	-4.415	-7.507	-12.608	
6	-10.575	-1.842	-2.178	0.478	6.702	3.912	
7	5.279	7.800	5.708	1.132	-1.216	5.382	
8	1.679	-2.138	5.238	-1.007	2.722	3.732	
9	2.595	5.720	-0.643	-0.039	-1.315	-0.525	
10	0.727	-7.937	-6.892	-8.737	-11.393	-27.537	
11	-3.663	-4.204	3.509	6.261	9.254	12.349	
12	-6.260	0.883	-14.828	4.856	0.880	12.270	
13	6.534	-0.858	-8.796	-0.627	-1.287	-1.789	
14	8.769	-3.621	0.799	-3.469	-0.565	-4.124	
15	-2.607	-6.937	1.408	-0.649	-5.814	-3.941	
16	0.120	0.011	0.500	5.110	0.000	0.000	

Fitting the Measurement Model to the Imputed Girl Data

Select the **Close all** option on the **Windows** menu to close all open windows and use the **New** option on the **File** menu of the root window to load the **New** dialog box. Select the **Path Diagram** option from the list box on the **New** dialog box to load the **Save As** dialog box.

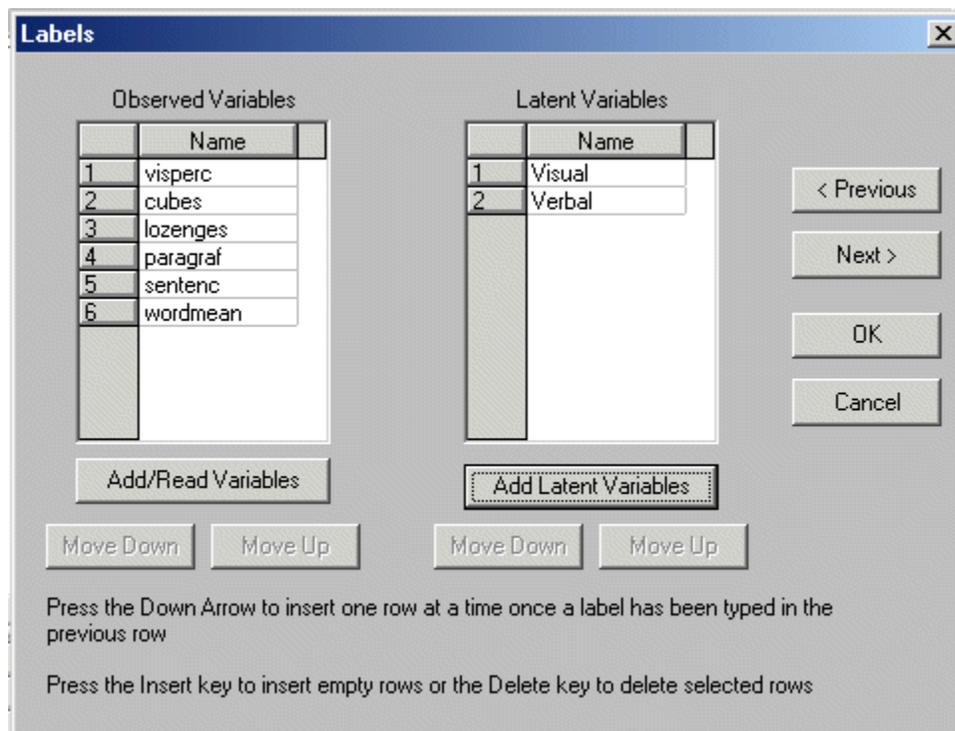
Enter **MGGIRLS.PTH** in the **File name** string field. Click on the **Save** push button to open an empty path diagram window. Select the **Title and Comments** option on the **Setup** menu to load the **Title and Comments** dialog to and enter A measurement model for visual perception and verbal ability of girls in the **Title** string field to produce the following **Title and Comments** dialog box.



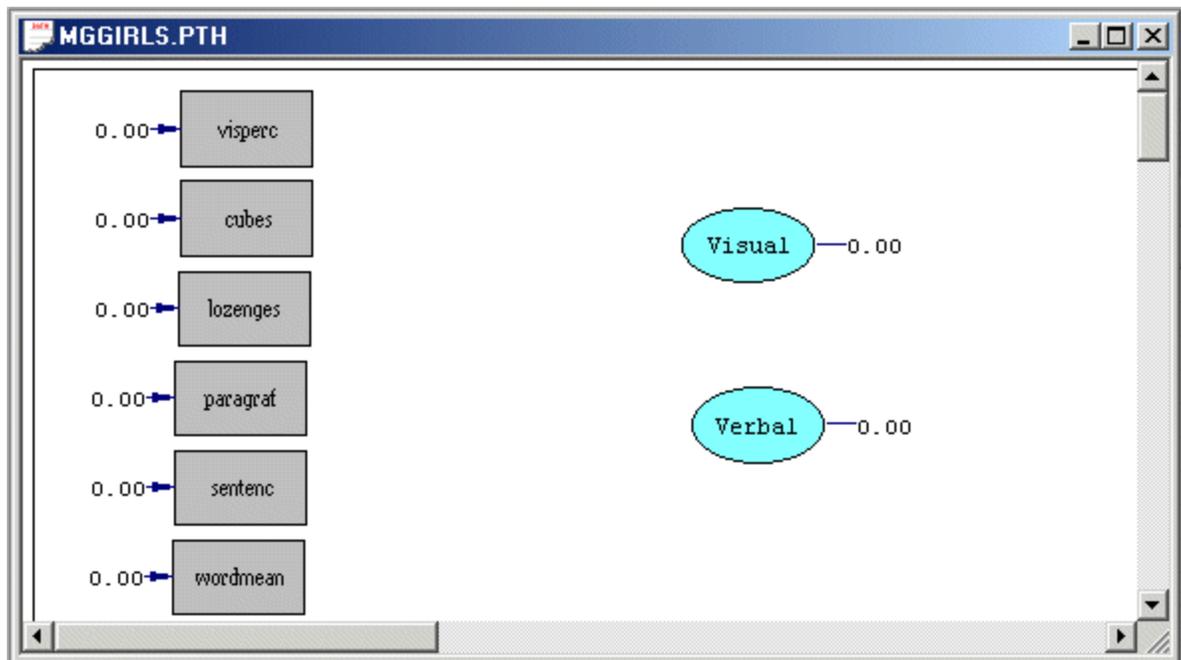
Click on the **Next** push button to load the **Group Names** dialog box. Click on the **Next** push button to load the **Labels** dialog box. Click on the **Add/Read Variables** push button to load the **Add/Read Variables** dialog box.

Select the **LISREL System File** option in the **Read from file:** dropdown list box. Click on the **Browse** push button and select the file **MGGIRLS_IMPUTED.LSF**. Click **OK** to return to the **Labels** dialog box.

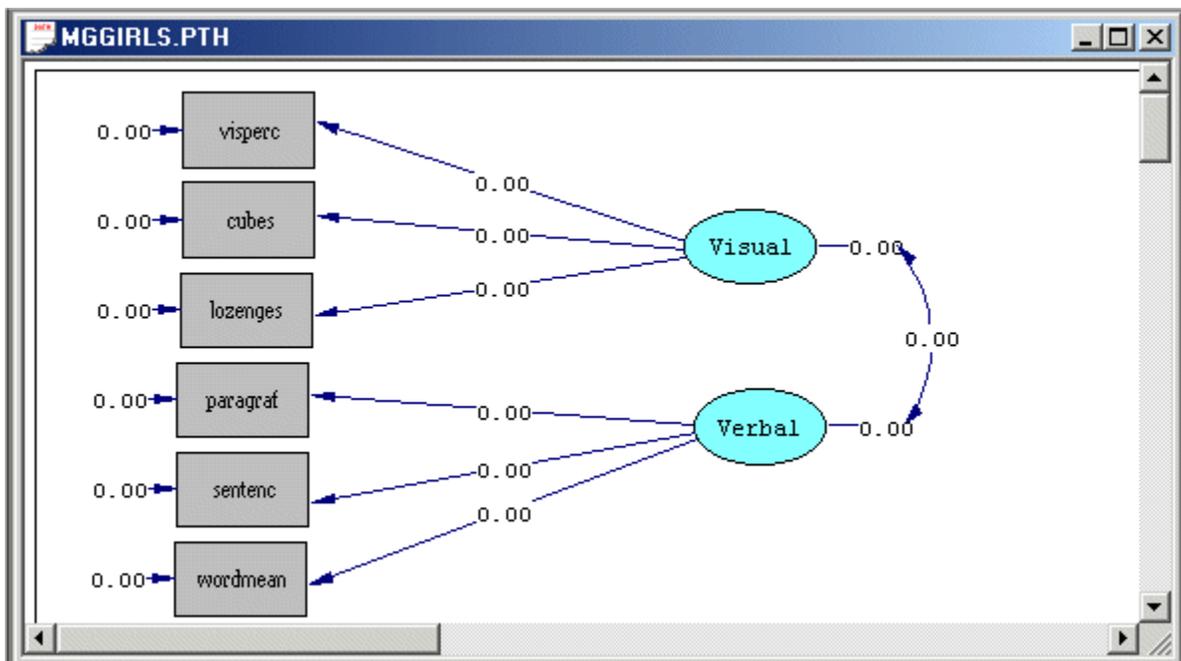
Click on the **Add Latent Variables** push button to load the **Add Variables** dialog box. Enter the label Visual for visual perception in the string field and click **OK** to return to the **Labels** dialog box. Repeat this to enter the label Verbal for verbal ability in the string field. Click **OK** to produce the following **Labels** dialog box.



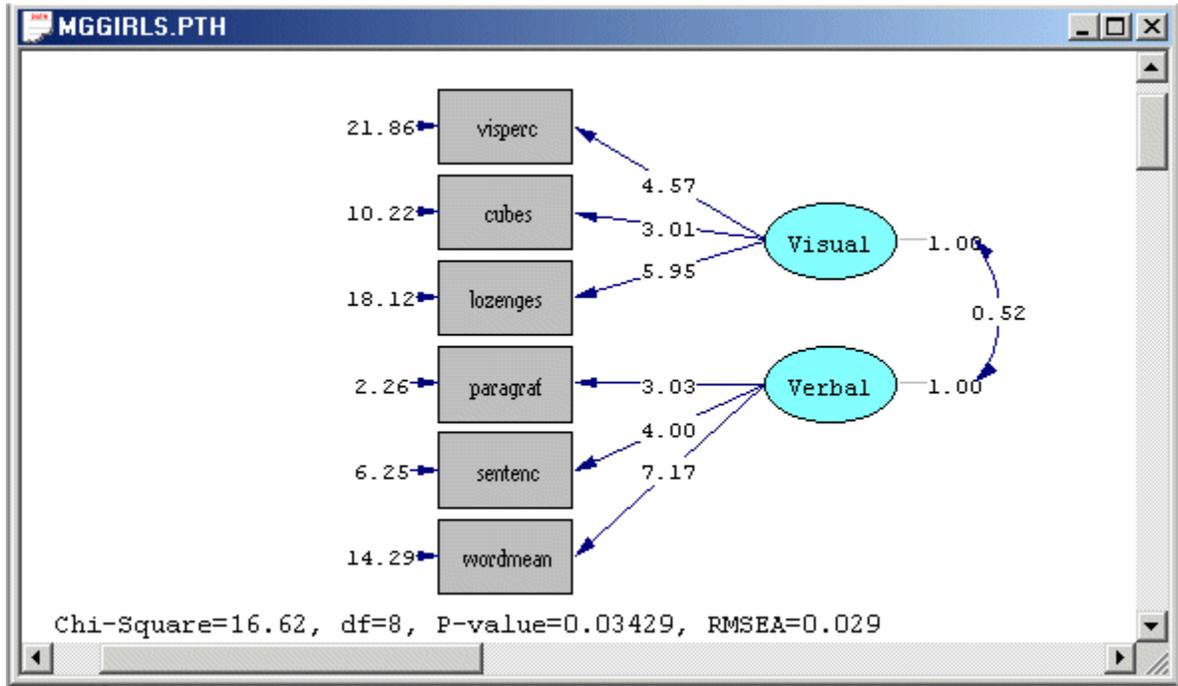
Click **OK** to return to the PTH window for **MGGIRLS.PTH**. Click, drag and drop the labels of the observed variables one at a time into the PTH window. Click, drag and drop the labels of the latent variables one at a time into the PTH window to produce the following PTH window.



Select the arrow icon on the drawing toolbar and Click and drag paths from Visual to visperc, cubes and lozenges respectively. Click and drag paths from Verbal to visperc, cubes and lozenges respectively. Next, select the two-headed arrow icon on the drawing toolbar and use it to click and drag a path from Visual to Verbal to produce the following PTH window.



Select the **Build SIMPLIS Syntax** option on the **Setup** menu to open the SIMPLIS project window. Click on the **Run LISREL** icon on the main toolbar to produce the following PTH window.



5. The FIML Method for Data with Missing Values

Close the PTH window for **MGGIRLS.PTH**. Close the output window for **MGGIRLS.OUT** to return to the SPJ file **MGGIRLS.SPJ**.

Use the **Save As** option on the **File** menu of the root window to load the **Save As** dialog box. Enter the name **MGGIRLS_FIML.SPJ** in the **File name** string field. Click on **Save** to open the SPJ window for **MGGIRLS_FIML.SPJ**.

Change the file name **MGGIRLS_IMPUTED.LSF** in the second line to **MGGIRLS.LSF**. Click on the **Run LISREL** icon on the main toolbar to produce the following PTH window.

