Structural Equation Modeling Workshop using Mplus

Michael D. Biderman, Ph.D.

The University of Tennessee at Chattanooga

Information from this workshop

All documents – pptx, mp4, pdf – associated with the workshop are available for download at

www.utc.edu/faculty/michael-biderman

Michael-Biderman@utc.edu

mdbiderman@gmail.com

Agenda

- Part I - Lecture
  - EFA using Mplus
  - CFA using Mplus
  - Structural Equation Models (SEM) using Mplus
- Part II – Lab work (Hands on exercise)
  - Analyzing a Structural Equation Model using Mplus)

Mplus

Mplus is one of several programs written to perform analyses necessary for Structural Equation Modeling

Among the most popular are **LISREL**, **EQS**, **Amos** (in SPSS), **Mplus**, **Lavaan** (in R), **Stata**, **CALIS** (in SAS). There are several others.

For more information about Mplus, see www.statmodel.com

Latest from the distributors of Mplus – Version 8, a workshop, and text

- **Mplus Version 8** is now available. August 17-18: 2-day **workshop at Johns Hopkins University**

(preceded by a 1-day course on our book ***Regression and Mediation Analysis using Mplus***).
The early bird fee is $25/day for students and postdocs and $75/day for others.
For details go to www.jhsph.edu/research/centers-and-institutes/johns-hopkins-center-for-prevention-and-early-intervention/Conferences/index

Exploratory Factor Analysis

- My take: EFA is a collection of techniques for identifying item clusters based on inter-item correlations.
- The goal is to identify groups of items which correlate more highly with each other than with other items in the total collection.
- Typically, the clusters result in a partitioning of the items.
- It is assumed that item clusters represent underlying characteristics which are called factors or latent variables.

Computational procedures of EFA

- Computer programs compute "predicted correlations" based on
  - 1) a specific number of factors and
  - 2) "guesses" concerning how much each item is influenced by each factor, called the factor **loadings**.
- The programs compare the predicted with the observed correlations.
- They then revise the "guesses" until the predicted correlations are as close as possible to the observed.
- Finally, the programs adjust or **rotate** the solution so that the correspondence between loadings and factors is as close as possible.
- The programs then report the solution and measures of goodness-of-fit – how well the predicted correlations correspond to the actual correlations

EFA Trivia

- Each factor influences ALL items in an EFA – all of the loadings are expected to be non-zero.
  - The items influenced the most by a factor are used to define that factor
- Overall goodness-of-fit is completely determined by the number of factors – more factors, better fit.
- The key is to find the number of factors that creates the best balance between theory, common sense, and goodness-of-fit.

Example EFA

HEXACO PI-R 100 Item Personality Questionnaire

Kibeom Lee, Ph.D., & Michael C. Ashton, Ph.D.

The HEXACO is a 100 item personality questionnaire – 96 items form 6 scales with 4 “filler” items.

It measures Honesty/humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, and Openness.

The E, X, A, C, and O scales are nearly identical to the Big Five personality dimensions. HEXACO adds a 6th- Honesty/Humility

HEXACO Facets

There are 16 items in each HEXACO scale or domain.

The items of each scale were written to form 4 groups, i.e., facets.

So, presumably, the questionnaire comprises 6 x 4 = 24 separate groups of items.

The focus here is on the Conscientiousness items and the 4 facets of that scale.

HEXACO Conscientiousness Facets

- **Org**anization
- 1 I clean my office or home quite frequently.
- 5 I plan ahead and organize things, to avoid scrambling at the last minute.
- 9 People often joke with me about the messiness of my room or desk.
- 13 When working, I sometimes have difficulties due to being disorganized.
- **Dili**gence
- 2 When working, I often set ambitious goals for myself.
- 6 I often push myself very hard when trying to achieve a goal.
- 10 Often when I set a goal, I end up quitting without having reached it.
- 14 I do only the minimum amount of work needed to get by.
- **Per**fectionism
- 3 I often check my work over repeatedly to find any mistakes.
- 7 When working on something, I don't pay much attention to small details.
- 11 I always try to be accurate in my work, even at the expense of time.
- 15 People often call me a perfectionist.
- **Pru**dence
- 4 I make decisions based on the feeling of the moment rather than on careful thought.
- 8 I make a lot of mistakes because I don't think before I act.
- 12 I don't allow my impulses to govern my behavior.
- 16 I prefer to do whatever comes to mind, rather than stick to a plan.

The HEXACO authors believed these items form 4 clusters. We'll evaluate the utility of that belief.

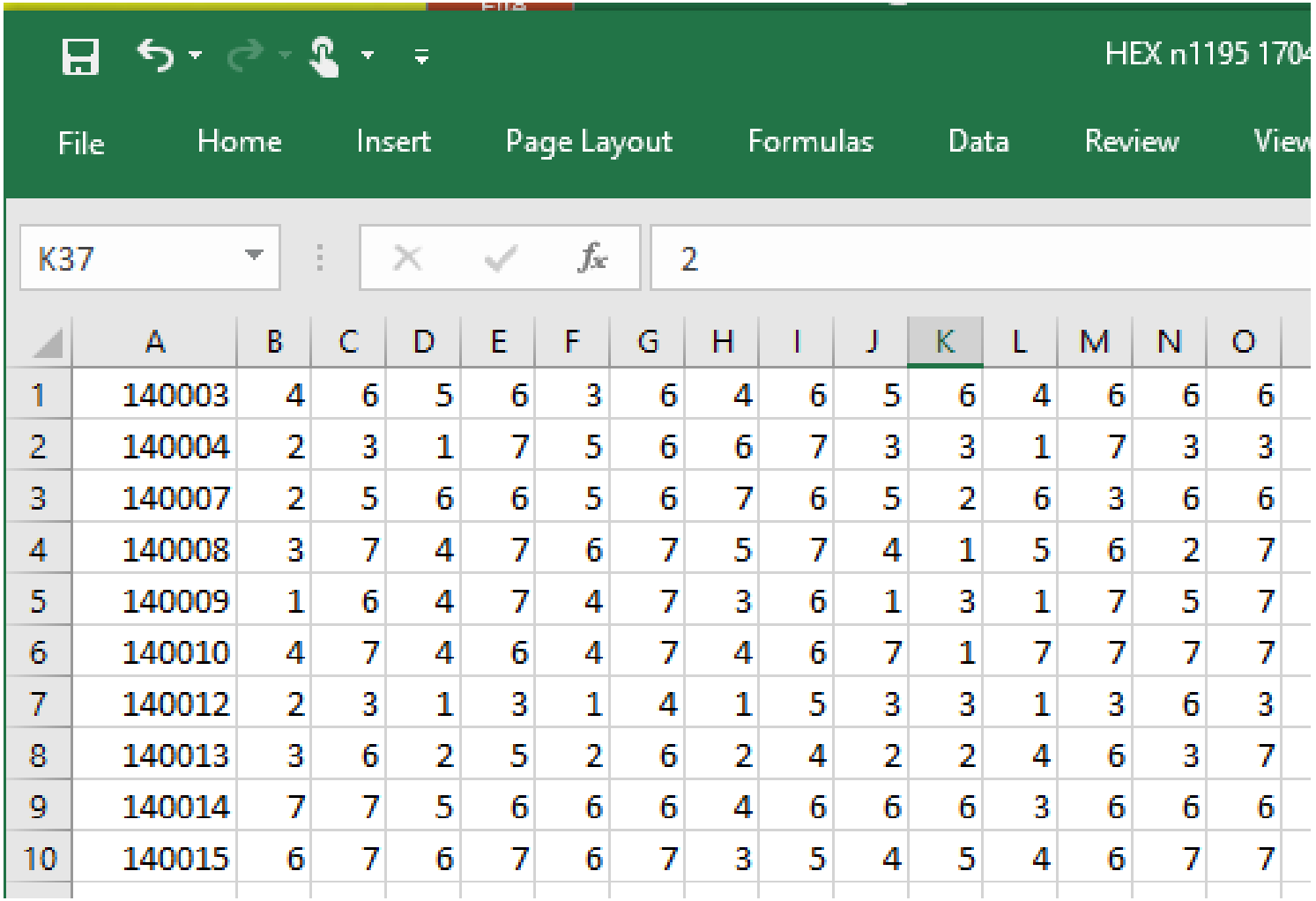
EFA Questions

- Do the responses to the HEXACO Conscientiousness items group into 4 clusters?
- Do the clusters that we find correspond to those proposed by the HEXACO creators?
- Since all items are Conscientiousness items, do all items represent the same characteristic?
- How should the data of the Conscientiousness scale be conceptualized – 4 separate facets or one dimension + something else?
-

The Data

- 1195 Undergraduates at a Southeastern University
- Responded online in Spring 2014 - Spring 2016
- Took NEO-FFI-3 and HEXACO-PI-R-100
- Respondents allowed gathering of GPA data
- Responses were on a 1 – 7 scale
- Negatively-worded items were reverse-scored

A peek at some of the data in Excel



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	140003	4	6	5	6	3	6	4	6	5	6	4	6	6	6
2	140004	2	3	1	7	5	6	6	7	3	3	1	7	3	3
3	140007	2	5	6	6	5	6	7	6	5	2	6	3	6	6
4	140008	3	7	4	7	6	7	5	7	4	1	5	6	2	7
5	140009	1	6	4	7	4	7	3	6	1	3	1	7	5	7
6	140010	4	7	4	6	4	7	4	6	7	1	7	7	7	7
7	140012	2	3	1	3	1	4	1	5	3	3	1	3	6	3
8	140013	3	6	2	5	2	6	2	4	2	2	4	6	3	7
9	140014	7	7	5	6	6	6	4	6	6	6	3	6	6	6
10	140015	6	7	6	7	6	7	3	5	4	5	4	6	7	7

HEXACO Conscientiousness Correlations

	hc1	hc5	hc9	hc13	hc2	hc6	hc10	hc14	hc3	hc7	hc11	hc15	hc4	hc8	hc12	hc16
hc1	1	.356	.533	.394	.273	.246	.199	.215	.223	.196	.195	.246	.162	.167	.133	.149
hc5	.356	1	.359	.425	.358	.402	.305	.333	.350	.285	.334	.350	.251	.332	.195	.343
hc9	.533	.359	1	.551	.183	.170	.274	.252	.126	.213	.152	.260	.225	.268	.141	.195
hc13	.394	.425	.551	1	.237	.240	.369	.397	.174	.324	.261	.312	.286	.348	.178	.300
hc2	.273	.358	.183	.237	1	.563	.364	.351	.354	.309	.348	.305	.125	.206	.166	.140
hc6	.246	.402	.170	.240	.563	1	.471	.410	.398	.318	.456	.317	.144	.182	.128	.145
hc10	.199	.305	.274	.369	.364	.471	1	.424	.238	.320	.339	.231	.209	.300	.103	.229
hc14	.215	.333	.252	.397	.351	.410	.424	1	.324	.348	.390	.288	.253	.288	.165	.223
hc3	.223	.350	.126	.174	.354	.398	.238	.324	1	.274	.360	.286	.145	.178	.151	.115
hc7	.196	.285	.213	.324	.309	.318	.320	.348	.274	1	.397	.357	.299	.358	.160	.217
hc11	.195	.334	.152	.261	.348	.456	.339	.390	.360	.397	1	.351	.205	.230	.217	.106
hc15	.246	.350	.260	.312	.305	.317	.231	.288	.286	.357	.351	1	.221	.212	.195	.232
hc4	.162	.251	.225	.286	.125	.144	.209	.253	.145	.299	.205	.221	1	.529	.395	.320
hc8	.167	.332	.268	.348	.206	.182	.300	.288	.178	.358	.230	.212	.529	1	.371	.328
hc12	.133	.195	.141	.178	.166	.128	.103	.165	.151	.160	.217	.195	.395	.371	1	.145
hc16	.149	.343	.195	.300	.140	.145	.229	.223	.115	.217	.106	.232	.320	.328	.145	1

Mplus

Basic Mplus Commands

title:	The Title of the analysis;
data:	The location of the data. ARGH!!;
variable:	Names of the variables;
usevariables are	Names of variables that are to be analyzed in this particular analysis;
analysis:	Specifications of this analysis.

Mplus Commands Example - EFA

```
title:              EAOM Mplus Workshop;

data:               FILE IS
'C:\Users\mdbid\Desktop\HEX GPA n1195 170409.csv';

variable:     names are
              bidid
              hh1 - hh16 hs1 - hs16
              hx1 - hx16 ha1 - ha16
              hc1 - hc16 ho1 - ho16 eosgpa;
usevariables are hc1-hc16;

analysis:     type = efa 1 4;
              rotation = geomin(Oblique);
```

Note that each command ends with a semicolon.

Mplus Commands Example - ESEM

```
title:              EAOM Mplus Workshop;

data:               FILE IS
'C:\Users\mdbid\Desktop\HEX GPA n1195 170409.csv';

variable:     names are
              bidid
              hh1 - hh16 hs1 - hs16
                    hx1 - hx16 ha1 - ha16
                    hc1 - hc16 ho1 - ho16 eosgpa;
usevariables are hc1-hc16;

analysis:     type = general;
              rotation = geomin(Oblique);

model:        f1 – f4 by hc1 – hc16 (*1);
```

Mplus has recently introduced an alternative EFA model, known as ESEM, for Exploratory Structural Equation Model.

EFA Goodness-of-fit – 1 Factor Solution

Chi-Square Test of Model Fit

Value	1603.350		
Degrees of Freedom	104		
P-Value	0.0000		P-value should be > .05

RMSEA (Root Mean Square Error Of Approximation)

Estimate		0.110	ARGH!	RMSEA should be <= .05
90 Percent C.I.	0.105	0.115		
Probability RMSEA <= .05		0.000		

CFI/TLI

CFI	0.727	ARGH!	CFI value should be >= .95
TLI	0.685	ARGH!	TLI value ahould be >= .90

SRMR (Standardized Root Mean Square Residual)

Value	0.075	ARGH!	SRMR value should be <= .05

Complete output is in the workshop handout.

Goodness-of-fit is not acceptable.

1 Factor Interpretation

A 1 factor solution is one in which all the items reflect the influence of one characteristic and no other characteristic.

The poor goodness-of-fit statistics suggest that responses to these items were influenced by more than characteristic.

EFA GOF – 2 & 3 Factor Solutions

2 Factor Solution

Chi-Square Test of Model Fit

Value	993.332
Degrees of Freedom	89
P-Value	**0.0000**

RMSEA (Root Mean Square Error Of Approximation)

Estimate		**0.092**
90 Percent C.I.	0.087	0.097
Probability RMSEA <= .05		0.000

CFI/TLI

CFI	**0.862**
TLI	**0.814**

SRMR (Standardized Root Mean Square Residual)

Value	**0.051**

3 Factor Solution

Chi-Square Test of Model Fit

Value	499.757
Degrees of Freedom	75
P-Value	**0.0000**

RMSEA (Root Mean Square Error Of Approximation)

Estimate		**0.069**
90 Percent C.I.	0.063	0.075
Probability RMSEA <= .05		0.000

CFI/TLI

CFI	**0.935**
TLI	**0.896**

SRMR (Standardized Root Mean Square Residual)

Value	**0.030**

Goodness-of-fit is better than that of the 1-factor solution, but neither is acceptable.

EFA GOF – 4 Factor Solution

```
Chi-Square Test of Model Fit

          Value                             256.465
          Degrees of Freedom                    62
          P-Value                           0.0000          P-value should be > .05

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.051          RMSEA value should be<= .05
          90 Percent C.I.              0.045  0.058
          Probability RMSEA <= .05           0.366

CFI/TLI

          CFI                                0.965          CFI value should be >= .95
          TLI                                0.932          TLI value ahouls be >= .90

SRMR (Standardized Root Mean Square Residual)

          Value                              0.023          SRMR value should be <= .05
```

So this looks like a solution that is pretty close to being OK.

Comments on 4 factor solution

- The Chi-square p-value is almost NEVER > .05 for large samples. Get used to it SEMers.
- If we had been "flying blind" in the analysis rather than targeting a 4-factor solution we'd have considered some rules for determining the number of factors – the scree test, parallel analysis, for example.

Key information in a solution

- Factor **loadings** – measures of the extent to which items are influenced by a factor
- We look for large loadings of the items on the factor that is supposed to have the greatest influence on and small loadings on all other items.
- Factor **correlations** – measures of the extent to which factors correlate with each other
- Sometimes researchers hope for factor correlations to be nearly zero; other times they hope for them to be non-zero

EFA 4 Factor Loadings

)

GEOMIN ROTATED LOADINGS (* significant at 5% level)

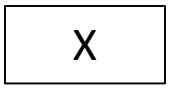
	1 Org	2 Dil	3 Per	4 Pru
HC1	0.727*	0.052*	-0.186*	-0.054
HC5	0.312*	0.335*	-0.027	0.152*
HC9	0.854*	-0.160*	0.067	-0.003
HC13	0.545*	0.033*	0.264*	0.116
HC2	0.045	0.679*	-0.041	-0.053
HC6	-0.042	0.839*	0.056	-0.122*
HC10	0.021	0.494*	0.422*	-0.019
HC14	0.049	0.464*	0.251*	0.091
HC3	0.025	0.538*	-0.150*	0.061
HC7	0.015	0.359*	0.108	0.259*
HC11	-0.041	0.585*	0.005	0.109*
HC15	0.182*	0.335*	-0.066	0.151*
HC4	-0.007	-0.049	-0.001	0.747*
HC8	0.009	0.021	0.109*	0.685*
HC12	0.002	0.036	-0.192*	0.558*
HC16	0.092*	0.039	0.141*	0.335*

Loadings of items on the appropriate facet have been red'd.

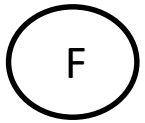
It appears that 8 items represent Factor 2 (Dil) in this solution and none represent Factor 3.

So, this solution, based on our data, suggest that Factors 2 and 3, in particular, Factor 3, need some work.

Path Diagrams



A rectangle represents an observed variable.

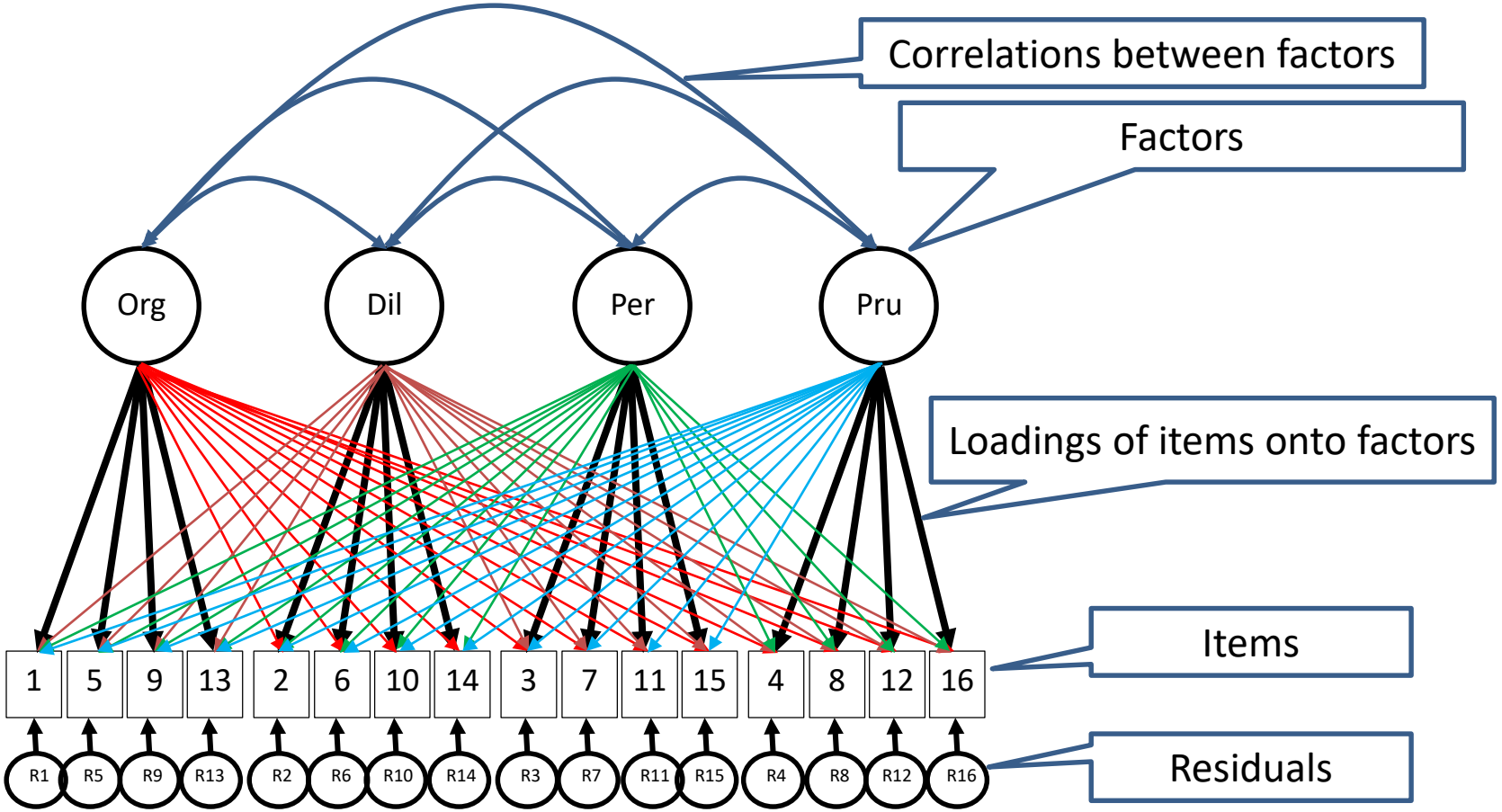


A circle/ellipse represents a factor

A single-headed arrow represents a loading.

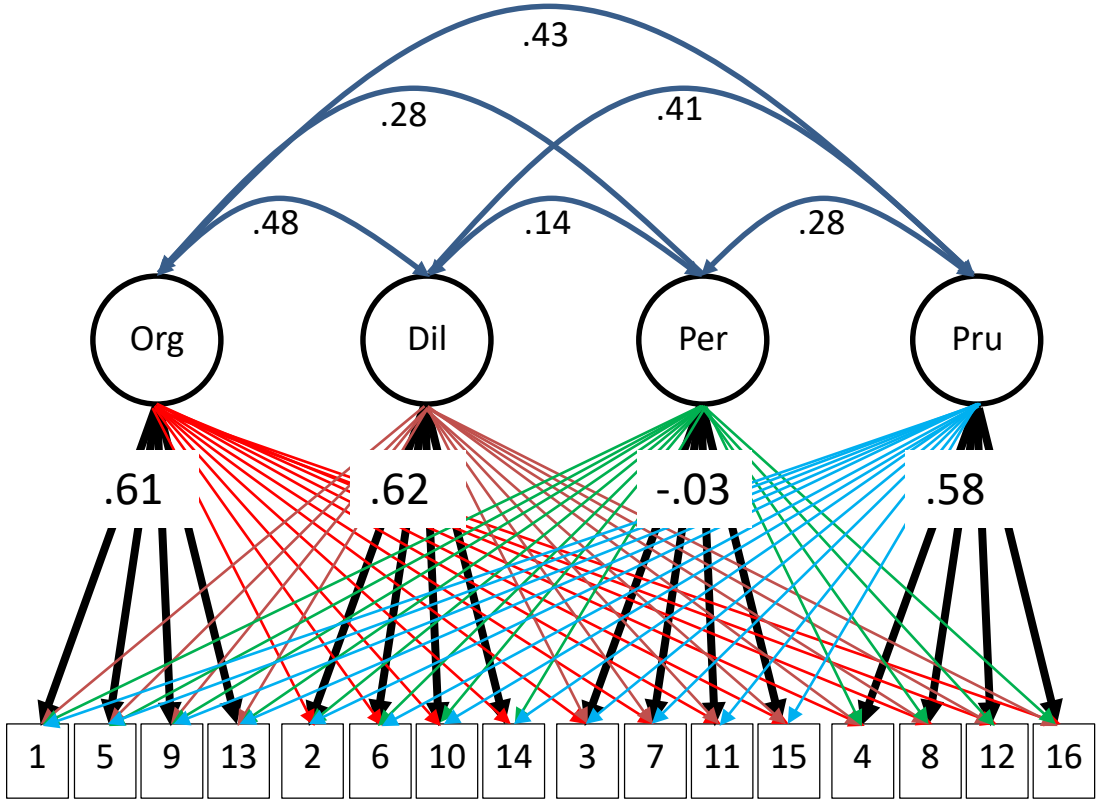
A double-headed arrow represents a correlation.

Path Diagram with Residuals



Residuals will not be represented in further path diagrams.

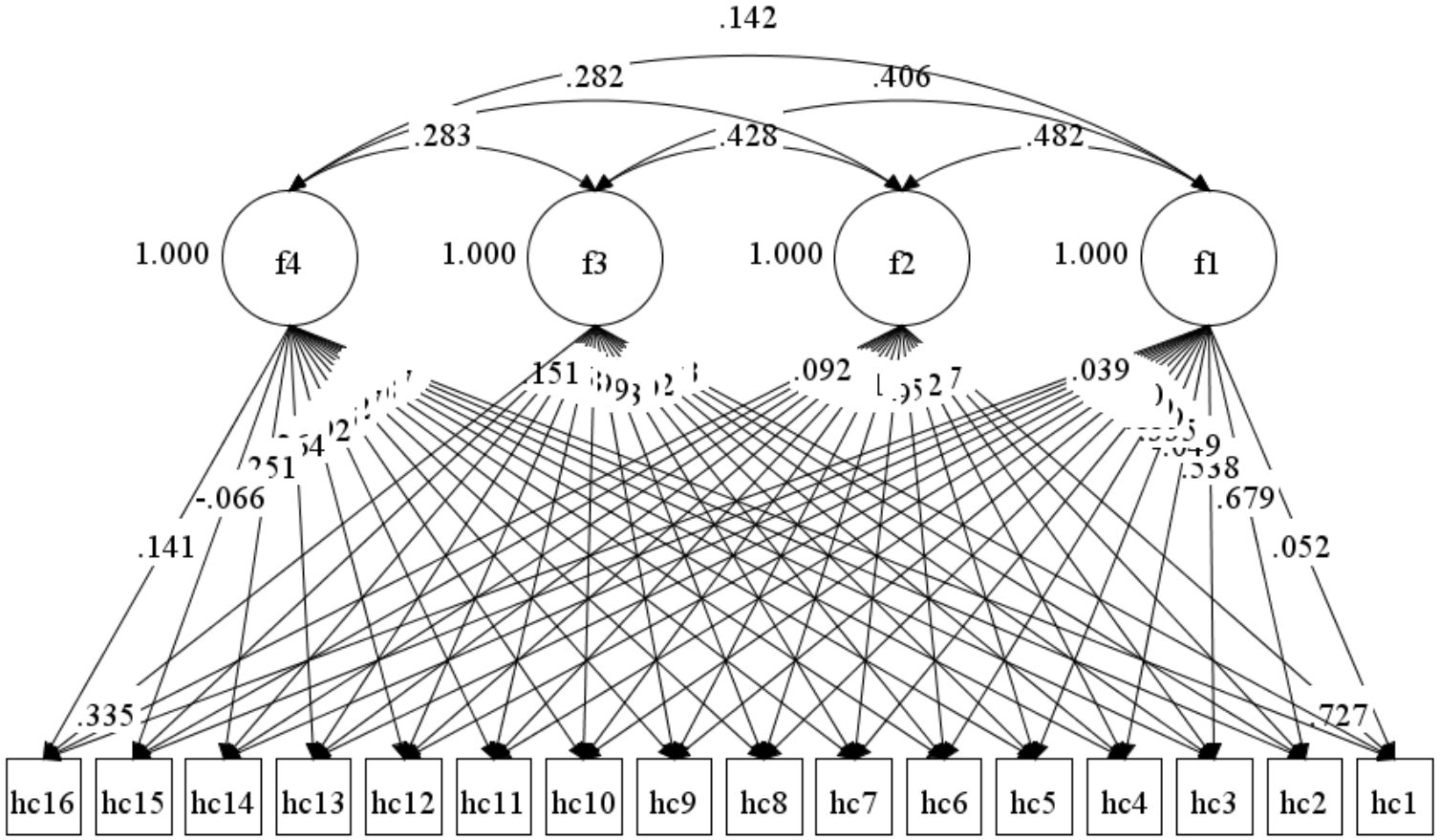
Path Diagram of 4-factor EFA Solution



- Colored paths are cross-loadings.
- With the exception of Per, the loading pattern seem to be what the authors intended.
- The factors are all positively correlated which is understandable since all are Conscientiousness facets.
- But where is the Conscientiousness factor?

Looking ahead - How can the characteristic common to ALL the items be represented?

Mplus's Diagram FYI



Scores based on the above EFA

- How would we use these results to guide actual measurement of persons.
- Most researchers and practitioners create summated scales.
- We could easily create three based on this – an Org facet scale, a Diligence facet scale, and a Prudence facet scale
- We might create an overall Conscientiousness scale as the sum of all items although there is no actual Conscientiousness factor.

The Missing Conscientiousness Factor

- The absence of a factor representing overall Conscientiousness is a problem..
- The only solution we've seen with a such a factor was the 1-factor solution, which fit terribly.
- So this particular EFA solution does not really represent what we think must be going on – multiple items all representing a single characteristic – Conscientiousness – AND also representing distinct facets.

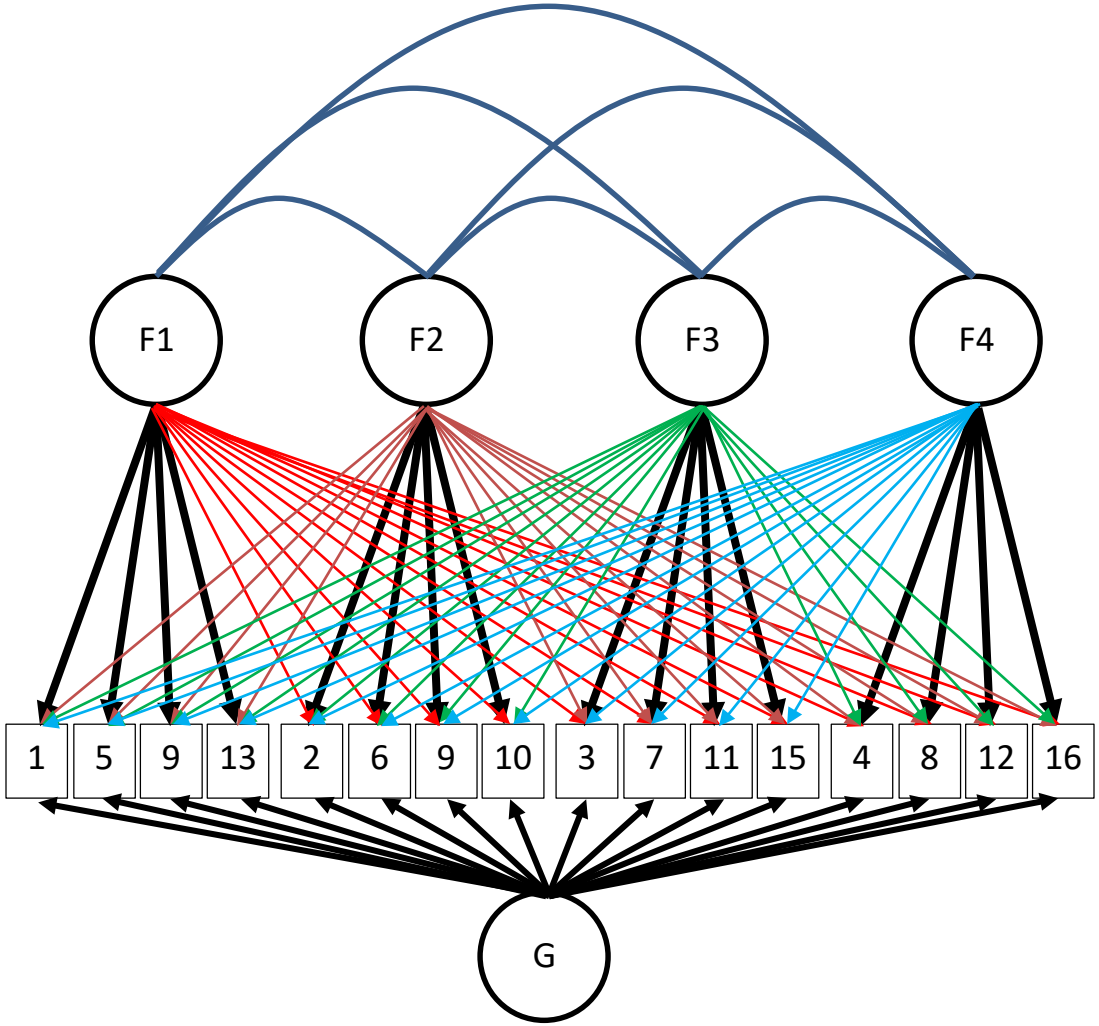
Need for a different type of solution

- Regular EFAs partition the items – into facets in this example.
- But the correlations between the facet factors suggest that there is a single characteristic common to all items.
- So what is needed is a hybrid solution – one that keeps the multi-factor partition but also acknowledges the what seems to be a common influence on all items.

The Bifactor Solution

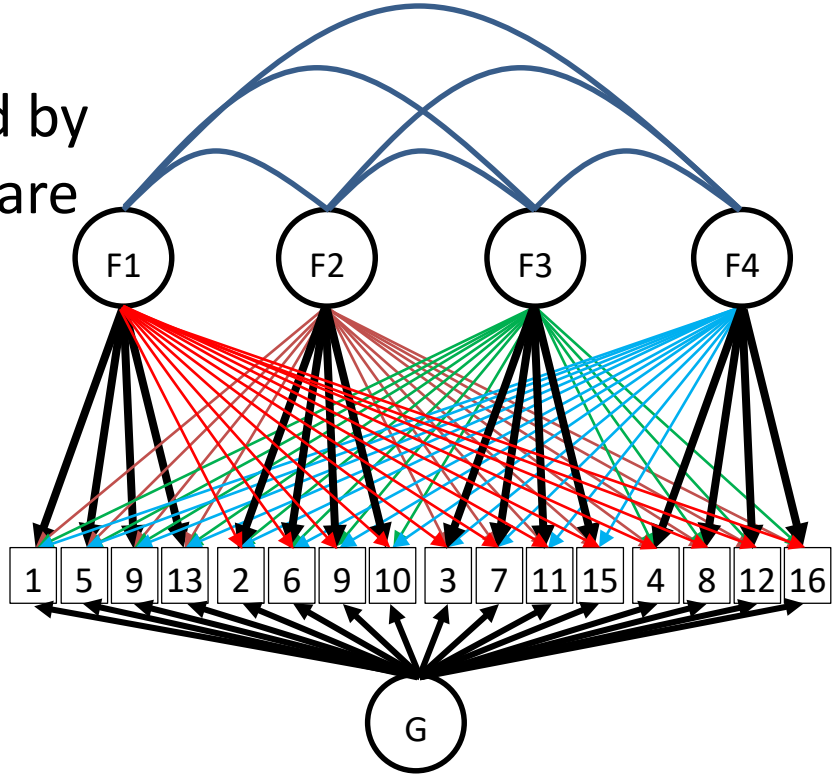
- Holzinger, K. J. & Swineford, F. in 1937 described a hybrid solution – the bifactor solution.
- A bifactor solution adds a single general factor to the regular EFA “partitioning” solution.
- It allows us to have our cake and eat it too.
- Use of bifactor models in CFA solutions has exploded in past 10 years
- A bifactor model is available as an EFA solution in Mplus – as an EFA only in Mplus I believe.
- .

Path Diagram of EFA Bifactor Solution



Bifactor Trivia - 1

Factors indicated by just some items are called Group Factors.



Correlations between group factors may be set to 0 or estimated. They're estimated here.

G is called the General Factor

For mathematical reasons, G must be estimated as being orthogonal to F1, F2, F3, and F4. So there are no double-headed arrows between G and the Fs.

Mplus EFA Bifactor Command File

```
title:              EAOM Mplus Workshop;

data:               FILE IS
'C:\Users\mdbid\Desktop\HEX GPA n1195 170409.csv';

variable:           names are
                    bidid
                    hh1 - hh16 hs1 - hs16
                hx1 - hx16 ha1 - ha16
                hc1 - hc16 ho1 - ho16 eosgpa ;
usevariables are hc1-hc16;

analysis:           type = efa 5 5;
                    rotation = bi-geomin(Oblique);
```

- 5 factors – G + 4 Facet factors.

Mplus ESEM Bifactor Command File

```
title:              EAOM Mplus Workshop;

data:               FILE IS
'C:\Users\mdbid\Desktop\HEX GPA n1195 170409.csv';

variable:     names are
              bidid
              hh1 - hh16 he1 - he16
              hx1 - hx16 ha1 - ha16
              hc1 - hc16 ho1 - ho16 eosgpa;
usevariables are hc1-hc16;

analysis:     type = general;
              rotation = bi-geomin(Oblique);

model:         f1 – f5 by hc1-hc16;
```

- 5 factors –4 Facet factors and a general factor.

EFA Bifactor Goodness-of-fit

```
Chi-Square Test of Model Fit

          Value                             164.650
          Degrees of Freedom                     50
          P-Value                           0.0000          P-value should be > .05

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.044 YES!     RMSEA should be <= .05
          90 Percent C.I.                0.036  0.051
          Probability RMSEA <= .05           0.911

CFI/TLI

          CFI                                0.979 YES!     CFI value should be >= .95
          TLI                                0.950 YES!     TLI value ahould be >= .90

SRMR (Standardized Root Mean Square Residual)

          Value                              0.018 YES!      SRMR should be <= .05
```

So this solution meets all but the chi-square criterion for goodness-of-fit.

Bifactor Model Loadings

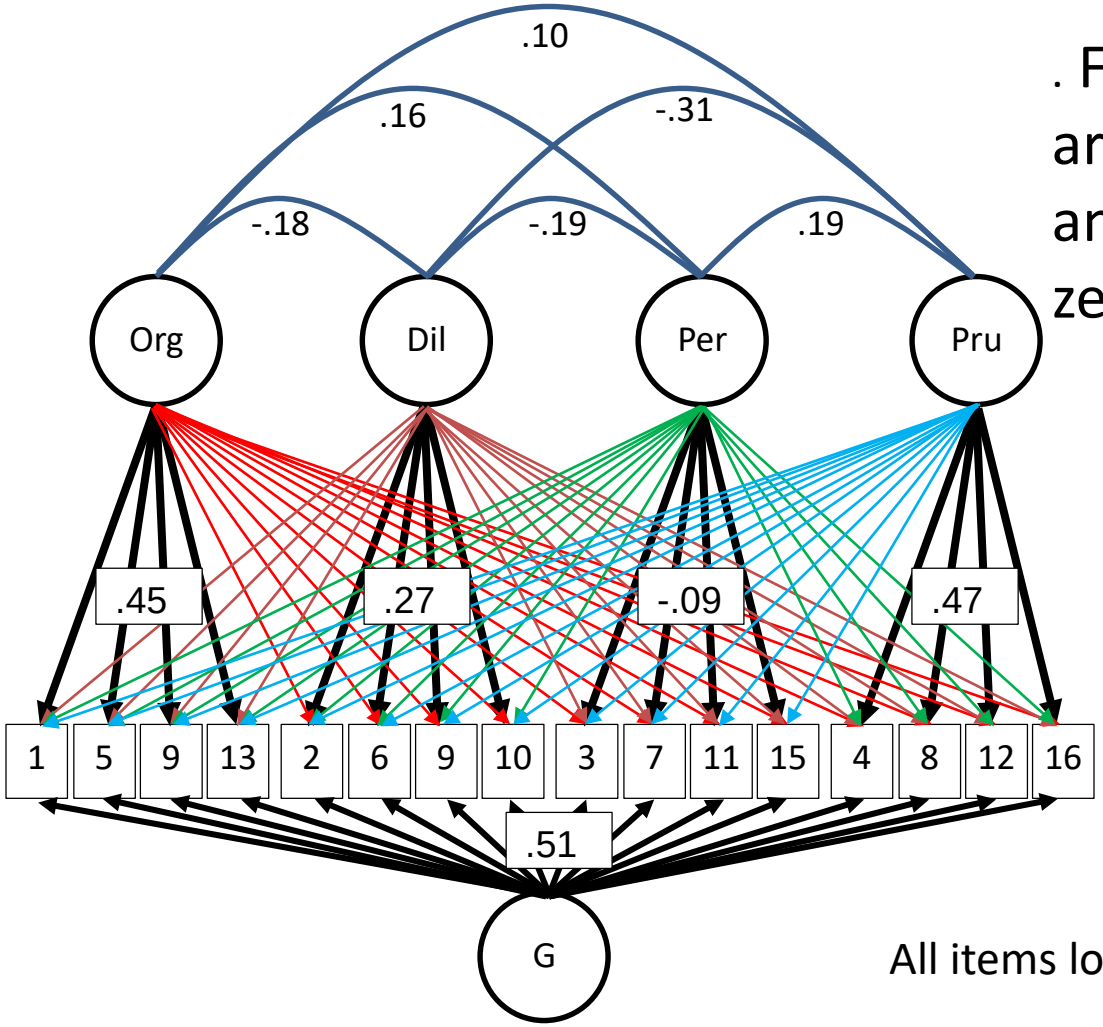
BI-GEOMIN ROTATED LOADINGS (* significant at 5% level)

	Gen	1 Org	2 Dil	3 Per	4 Pru
HC1	0.421*	0.567*	0.051	-0.205*	-0.020
HC5	0.600*	0.196*	0.056	-0.072	0.084*
HC9	0.425*	0.668*	-0.046	0.039	0.010
HC13	0.564*	0.372*	-0.139	0.215	0.003
HC2	0.583*	0.021	0.337*	-0.074	-0.003
HC6	0.670*	-0.049	0.465*	0.016	-0.052
HC10	0.566*	0.013	0.265*	0.394*	0.009
HC14	0.614*	-0.039	0.029	0.193*	-0.019
HC3	0.521*	-0.064	0.062	-0.191*	-0.030
HC7	0.583*	-0.111*	-0.135*	0.045	0.060
HC11	0.632*	-0.166*	-0.035	-0.066	-0.064
HC15	0.554*	0.022	-0.154*	-0.142	-0.034
HC4	0.393*	-0.010	-0.032	-0.021	0.604*
HC8	0.459*	0.004	0.013	0.087*	0.576*
HC12	0.305*	-0.009	0.000	-0.199*	0.454*
HC16	0.343*	0.056	-0.033	0.109	0.245*

ALL items loaded significantly on the general factor. It gets my vote as the Conscientiousness Factor.

The Per items don't load on any group factor. They just load, quite highly, on the Conscientiousness factor.

EFA Bifactor Solution Path Diagram



. Factor correlations are fairly close to and bounce around zero

Factor 3 is still a wash.

All items load on G, as we expected.

Comments on EFA Bifactor Solution

- The bifactor solution reflects the fact that the response to each item is determined by two influences – Conscientiousness and the facet content of that item.
- The bifactor model assumes that facet factors are independent of the general factor.
- This means that a person high in Conscientiousness is equally likely to be high or low in Organization
- The implications of the independence of the general factor and group factors is something that needs further investigation

Measuring Bifactor Model Factors

- The only way I know to measure the general characteristics and the facet characteristics **in a way that preserves the relationships in the model** is to use factor scores.
- Mplus will create factor scores and put them in a file. You must use the ESEM commands.

Mplus ESEM commands for EFA factor scores

```
title:                  EAOM Mplus Workshop;
data:                   FILE IS
'C:\Users\mdbid\Desktop\HEX GPA n1195 170409.csv';
variable:       names are
                bidid
                hh1 - hh16 he1 - he16
                hx1 - hx16 ha1 - ha16
                hc1 - hc16 ho1 - ho16 eosgpa;
usevariables are hc1-hc16;
analysis:       type = general;
                rotation = bi-geomin(Oblique);
model:           f1 - f5 by hc1 - hc16 (*1);
savedata: file is 'C:\Users\mdbid\Desktop\ESEM Bifactor 5 Factor
Scores.txt';
save=fscores;
```

The .txt file can be opened using Excel or SPSS.

Confirmatory Factor Analysis

- Factor analysis in which some loadings are fixed, typically at 0.

- Why perform a CFA?
  - Models with fewer loadings are easier to conceptualize
  - The mathematics of CFAs is much simpler than that of EFAs

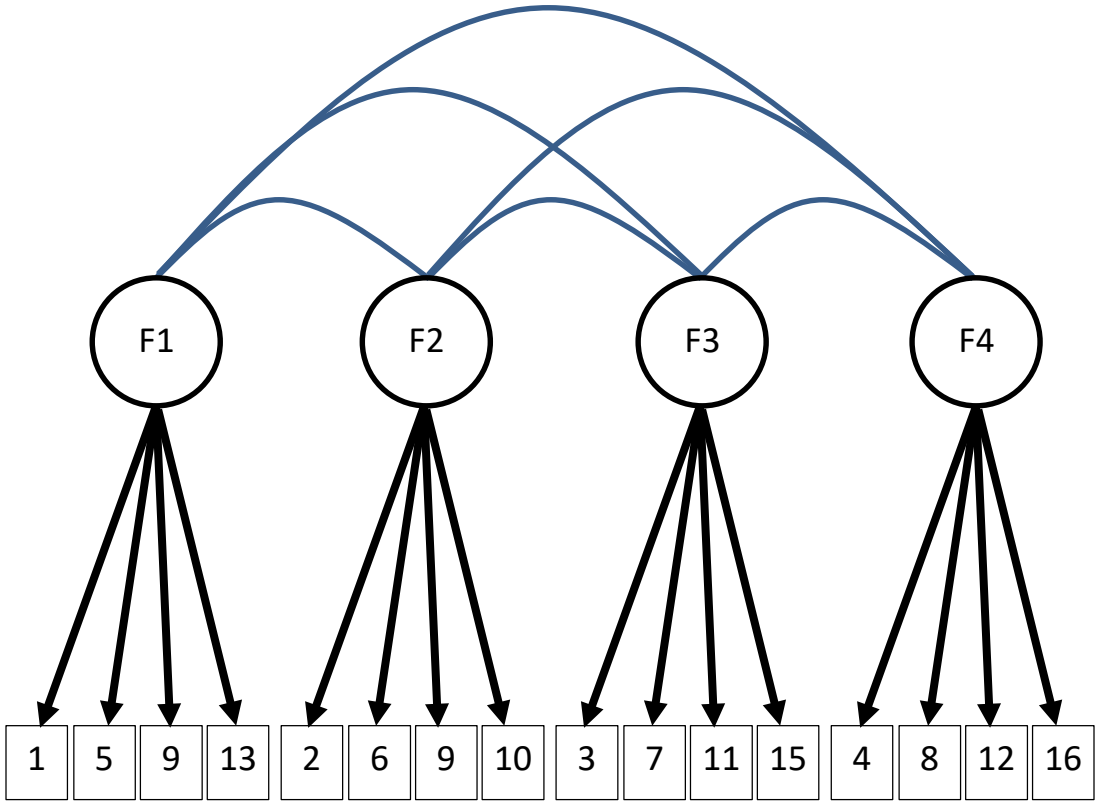
- When are CFAs conducted?
  - Sometimes a CFA is the only analysis considered
  - Sometimes CFAs are created based on patterns of loadings found in EFAs
    - Multiple EFAs of the same questionnaire have resulted in near zero loadings for many items – so why not fix them at 0?
    - A CFA is then tested in which the loadings are fixed at the values identified in the EFAs.

Suggestions for CFA from our EFA

	Gen	1 Org	2 Dil	3 Per	4 Pru
HC1	0.421*	0.567*	0	0	0
HC5	0.600*	0.196*	0	0	0
HC9	0.425*	0.668*	0	0	0
HC13	0.564*	0.372*	0	0	0
HC2	0.583*	0	0.337*	0	0
HC6	0.670*	0	0.465*	0	0
HC10	0.566*	0	0.265*	0	0
HC14	0.614*	0	0.029	0	0
HC3	0.521*	0	0	-0.191*	0
HC7	0.583*	0	0	0.045	0
HC11	0.632*	0	0	-0.066	0
HC15	0.554*	0	0	-0.142	0
HC4	0.393*	0	0	0	0.604*
HC8	0.459*	0	0	0	0.576*
HC12	0.305*	0	0	0	0.454*
HC16	0.343*	0	0	0	0.245*

- For our data, the EFA suggests the pattern of loadings red'd here, with the remaining loadings fixed at 0.
- I'll start with a simple Correlated Factors CFA, without a general factor.
- I'll estimate Factor 3 (Per) even though the loadings don't warrant it.

CFA Path Diagram



Note that the paths representing cross-loadings are gone – fixed at 0.

Mplus 4-factor CFA Commands

```
title:            EAOM Mplus Workshop;
data:             FILE IS
'C:\Users\mdbid\Desktop\HEX GPA n1195 170409.csv';
variable:         names are
                  bidid
                  hh1 - hh16 hs1 - hs16
                  hx1 - hx16 ha1 - ha16
                  hc1 - hc16 ho1 - ho16 eosgpa ;
usevariables are hc1-hc16;

analysis:   type = general;

model:      f1 by hc1 hc5 hc9  hc13;
            f2 by hc2 hc6 hc10 hc14;
            f3 by hc3 hc7 hc11 hc15;
            f4 by hc4 hc8 hc12 hc16;

Output:     STDYX;                  ! to get standardized estimates;
```

As a default Mplus estimates factor correlations.

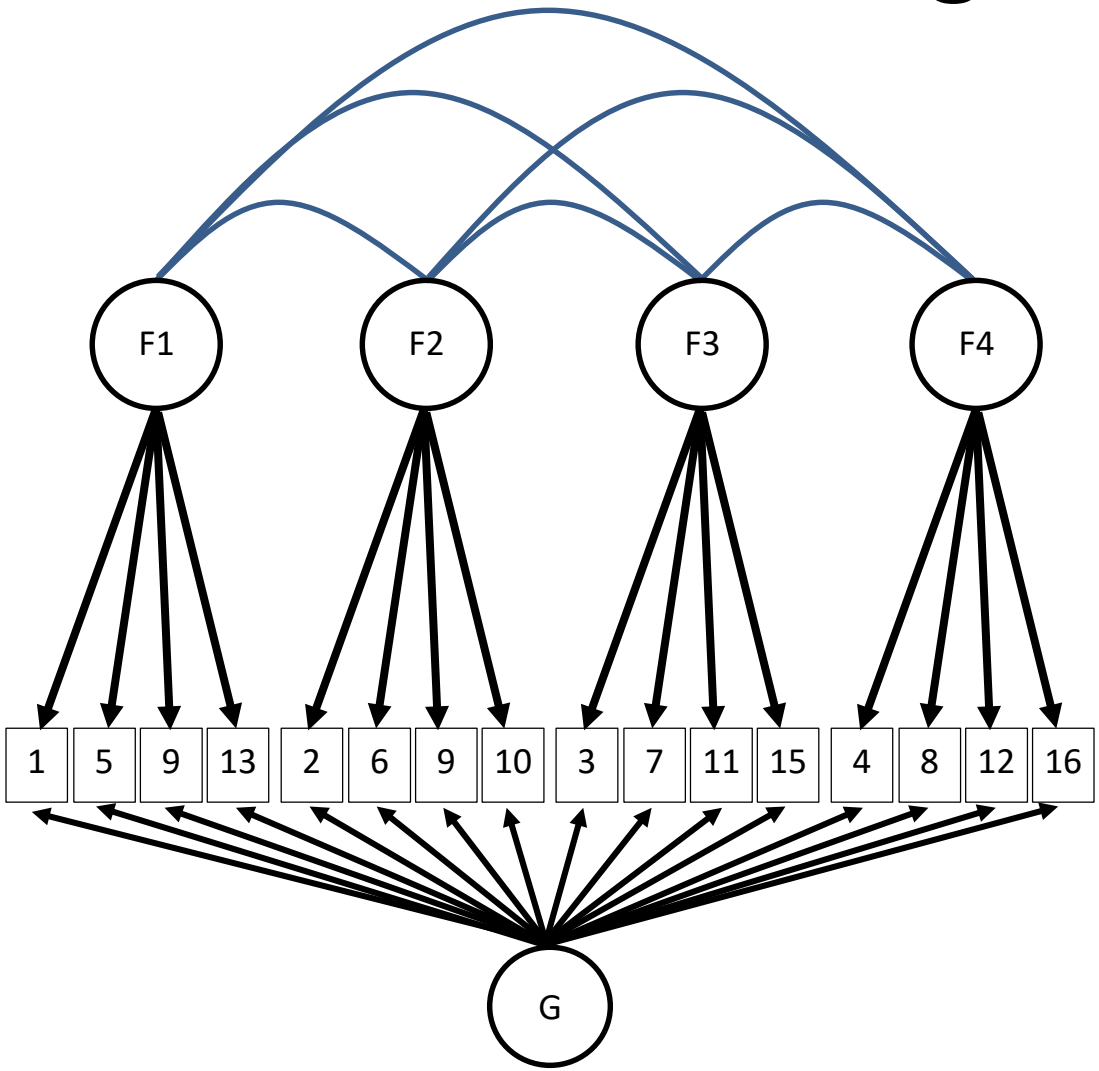
CFA Goodness-of-fit

```
Chi-Square Test of Model Fit

          Value                             740.889
          Degrees of Freedom                    98
          P-Value                           0.0000          P-value should be > .05

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.074          RMSEA value should be<= .05
          90 Percent C.I.                0.069  0.079
          Probability RMSEA <= .05           0.000

CFI/TLI

          CFI                                0.883          CFI value should be >= .95
          TLI                                0.857          TLI value ahouls be >= .90

SRMR (Standardized Root Mean Square Residual)

          Value                              0.056          SRMR value should be <= .05
```

The fit of this model is not so good.

Let's try a Bifactor CFA

Bifactor CFA Path Diagram



Mplus Bifactor CFA Commands

```
title:     EAOM Mplus Workshop;

data:       FILE IS 'C:\Users\mdbid\Desktop\HEX GPA n1195 170409.csv';

variable:    names are bidid hh1 - hh16 he1 - he16 hx1 - hx16 ha1 - ha16
                       hc1 - hc16 ho1 - ho16 eosgpa;
usevariables are hc1-hc16;

analysis:   Type = general;
model:      f1 by hc1 hc5  hc9 hc13;
            f2 by hc2 hc6 hc10 hc14;
            f3 by hc3 hc7 hc11 hc15;
            f4 by hc4 hc8 hc12 hc16;
            g by hc1 - hc16;
            g with f1-f4@0; !Fixes g~f1, g~f2, g~f3, g~f4 correlations = 0;

output:      STDYX;          ! To get standardized estimates;
```

Bifactor CFA Goodness-of-fit

```
Chi-Square Test of Model Fit

          Value                             371.104
          Degrees of Freedom                    82
          P-Value                           0.0000          P-value should be > .05

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.054  ???     RMSEA value should be<= .05
          90 Percent C.I.              0.049  0.060
          Probability RMSEA <= .05           0.099

CFI/TLI

          CFI                                0.947  ???     CFI value should be >= .95
          TLI                                0.923  YES     TLI value ahouls be >= .90

SRMR (Standardized Root Mean Square Residual)

          Value                              0.032  YES     SRMR value should be <= .05
```

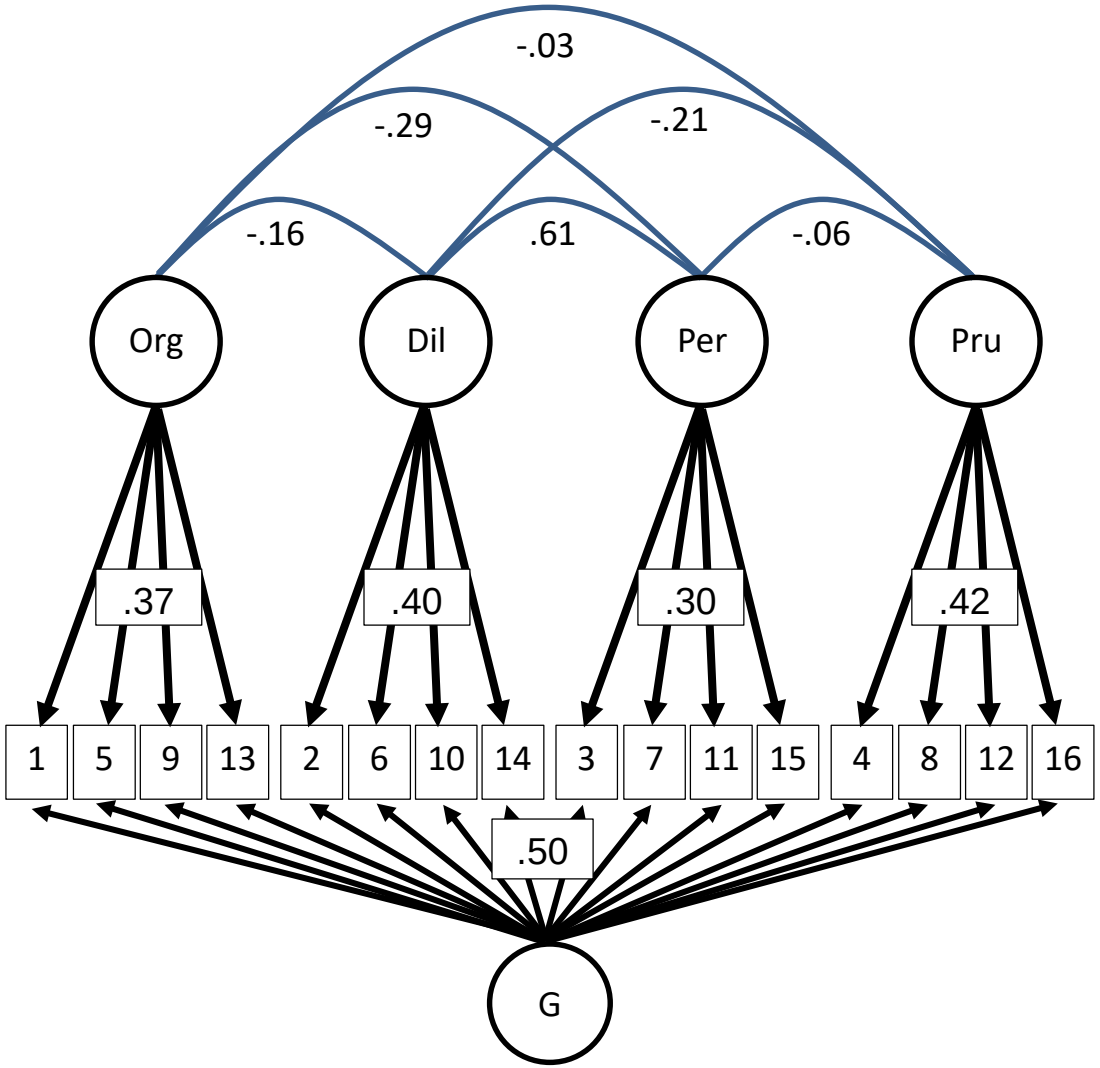
The fit of this model is better than the correlated factors CFA but not as good as the Bifactor EFA. We miss the cross-loadings.

CFA Bifactor Estimates

	G	1 Org	2 Dil	3 Per	4 Pru
HX1	0.455*	0.378*	0	0	0
HX5	0.649*	0.034	0	0	0
HX9	0.514*	0.788*	0	0	0
HX13	0.641*	0.283*	0	0	0
HX2	0.481*	0	0.453*	0	0
HX6	0.508*	0	0.701*	0	0
HX10	0.538*	0	0.271*	0	0
HX14	0.583*	0	0.184*	0	0
HX3	0.420*	0	0	0.368*	0
HX7	0.539*	0	0	0.207*	0
HX11	0.501*	0	0	0.469*	0
HX15	0.513*	0	0	0.171*	0
HX4	0.445*	0	0	0	0.617*
HX8	0.529*	0	0	0	0.487*
HX12	0.314*	0	0	0	0.402*
HX16	0.422*	0	0	0	0.196*

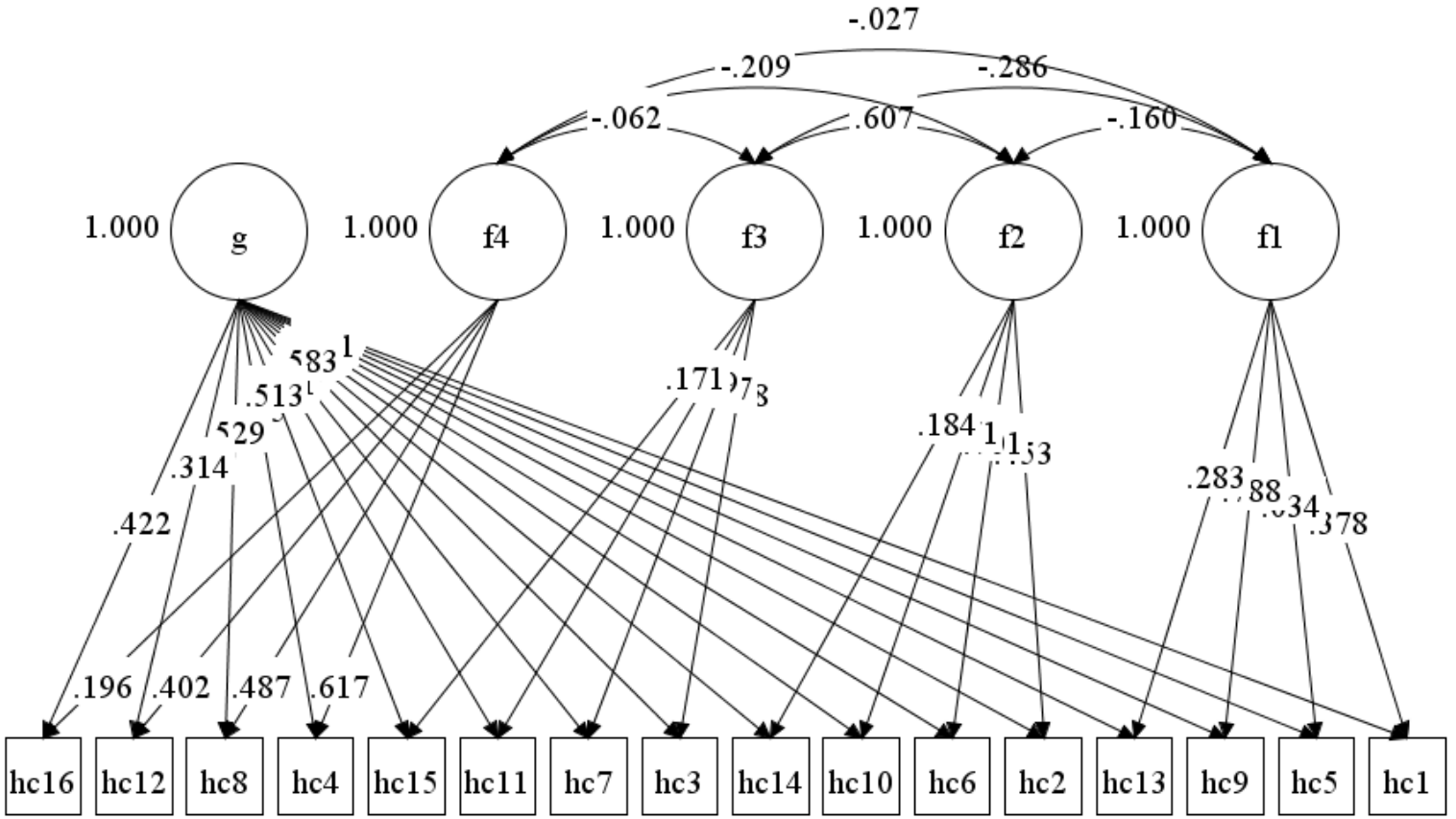
These estimates are about what we would expect, based on the previous EFAs.

Bifactor CFA Solution Path Diagram

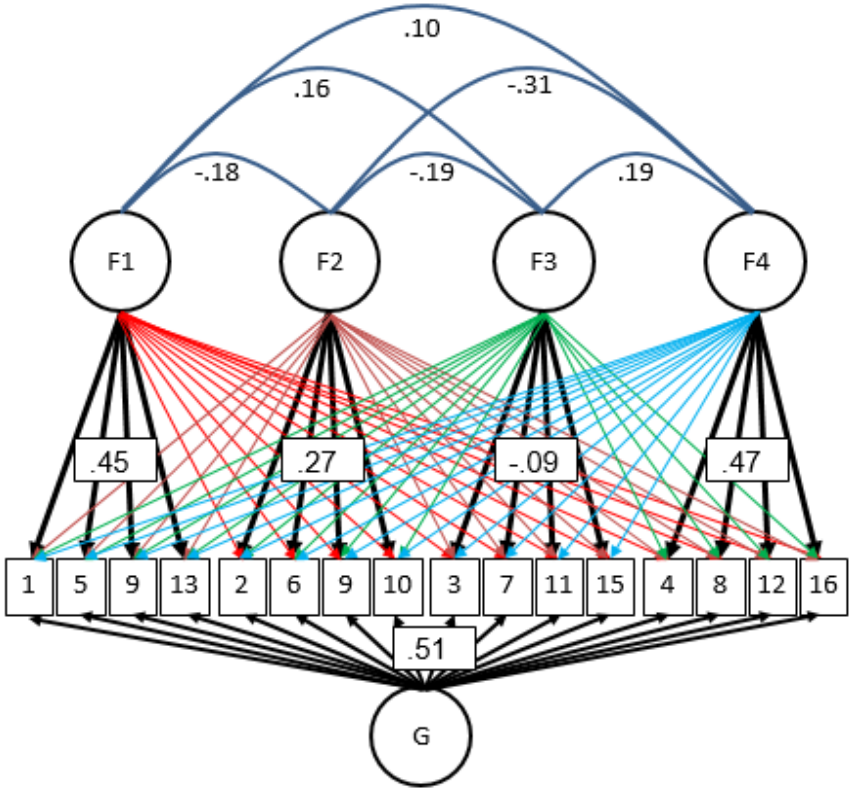


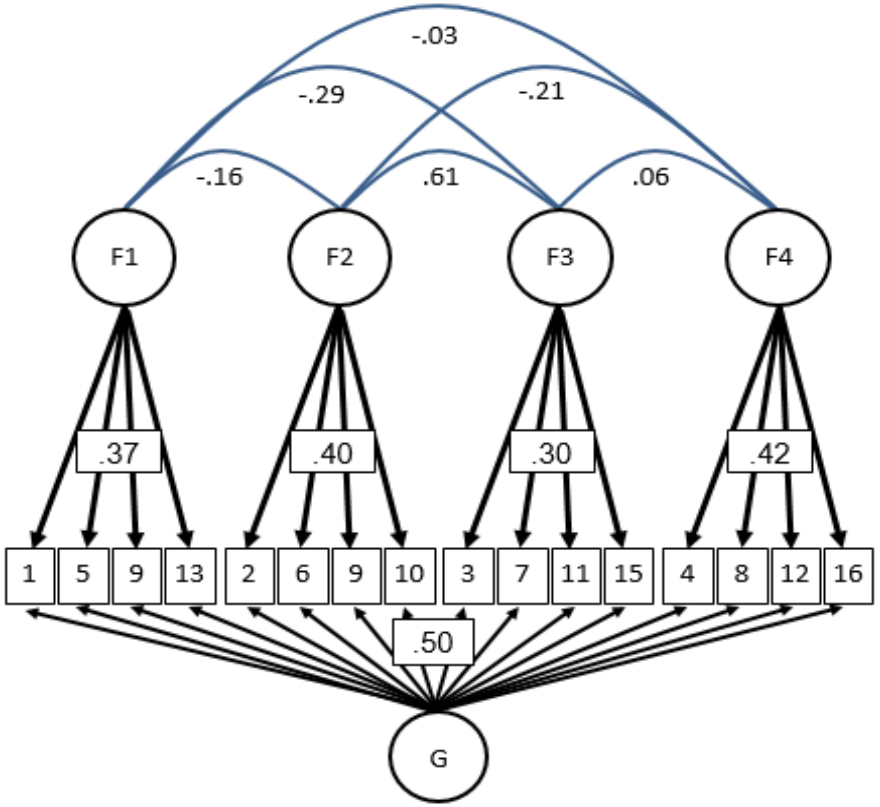
Note the large correlation between F2 (Dil) and F3 (Per). This follows previous results suggesting that the Dil and Per items are not very dissimilar.

Mplus's Diagram



EFA vs CFA Bifactor Solutions





The items presumed to define F3 (Per) essentially belong to G and the other factors, especially F2 (Dil)

Conclusions regarding HEXACO Conscientiousness

1. There is clearly a general Conscientiousness factor influencing responses to all the items – so it was worthwhile estimating a bifactor model – EFA or CFA.
2. In addition to Conscientiousness, there is clearly a tendency to be **Organized** that influenced the responses to items 1,5, 9, and 13 (Facet 1) in both EFA and CFA models.
3. In addition to Conscientiousness, there is clearly a tendency to be **Prudent** that influenced the responses to items 4, 8, 12, and 16 (Facet 4), again in both EFA and CFA models.
4. In addition to Conscientiousness, there was a weak tendency to be **Diligent** that influenced the responses to items 2,6,10, and 14 (Facet 2).
5. There is not strong evidence for a tendency to be **Perfectionist** influencing responses to items 3, 7, 11, and 15 (Facet 3).

Which is Preferred – EFA or CFA?

- Kline, R. B. (2016) *Principles and Practice of Structural Equation Modeling.* New York: Guilford.

Cross-loadings may account for so much variance that constraining them to zero may be too conservative.

". . . The more restrictive CFA may not fit the data."

". . . CFA is not inherently superior to EFA."

EFA or CFA? – 2

- Morin, Arens, Marsh (2016)
- “ . . . studies showed that EFA usually results in more exact estimates of the true population values for the latent factor correlations than CFA . . .”

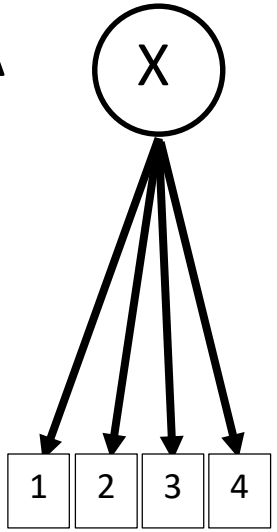
Structural Equation Models

- SEM refers to a collection of analytic techniques for testing **relationships** between variables some of which have been defined by CFAs.
- Here I will focus on analysis of regression relationships between variables, one or more of which is defined within a CFA
- The CFA defining a latent variable used within a SEM is called the **Measurement Model**.
- The relationship between X and Y specified in the SEM is called the **Structural Model**.

Measurement Models

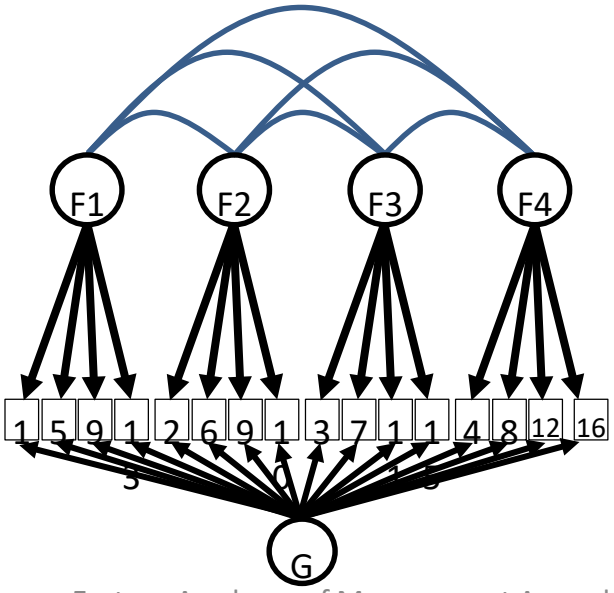
- Simplest measurement model,

  a single observed variable: X
- The next simplest type of measurement model is a simple CFA



Measurement Models - 2

A multifactor CFA, such as our CFA of the HEXACO Conscientiousness items, could be a measurement model for an SEM with each factor being an X variable or a Y variable



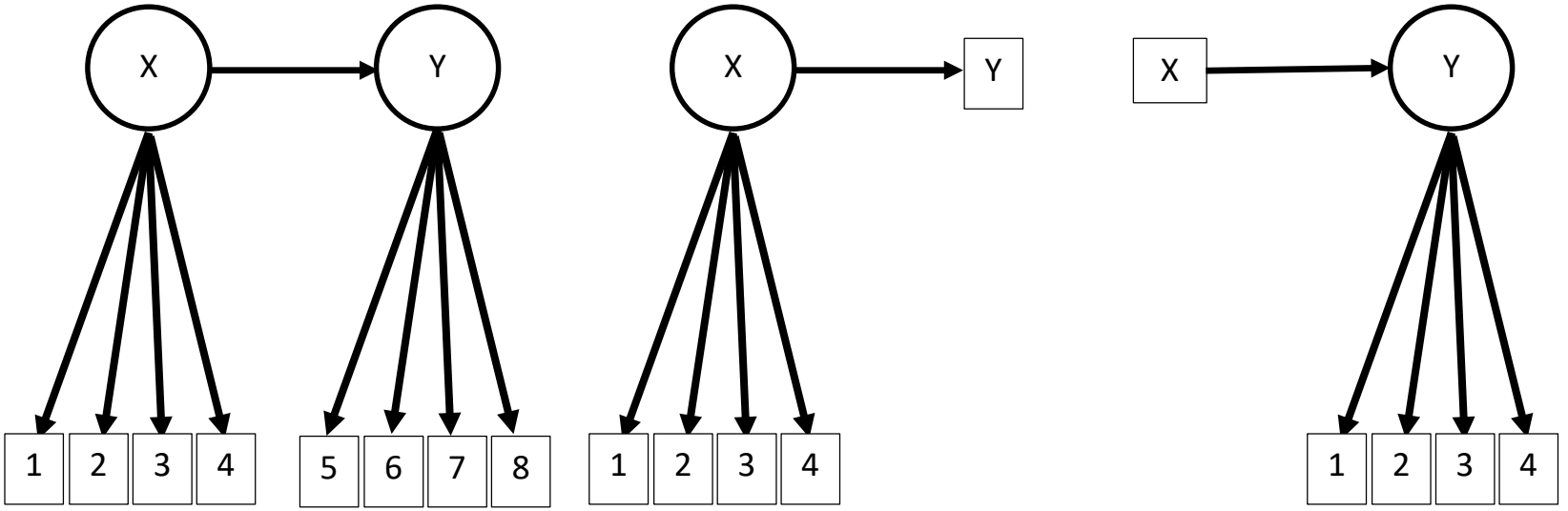
Structural Models

- A structural model is a specification of the relationship between X and Y, where either X or Y or both are latent variables, defined by measurement models.

- Usually: 

Three Example SEMs

- Latent X -> Latent Y Latent X -> Observed Y Observed X – Latent Y



The X or Y measurement model can be a CFA.

The X or Y measurement model can be an observed variable.

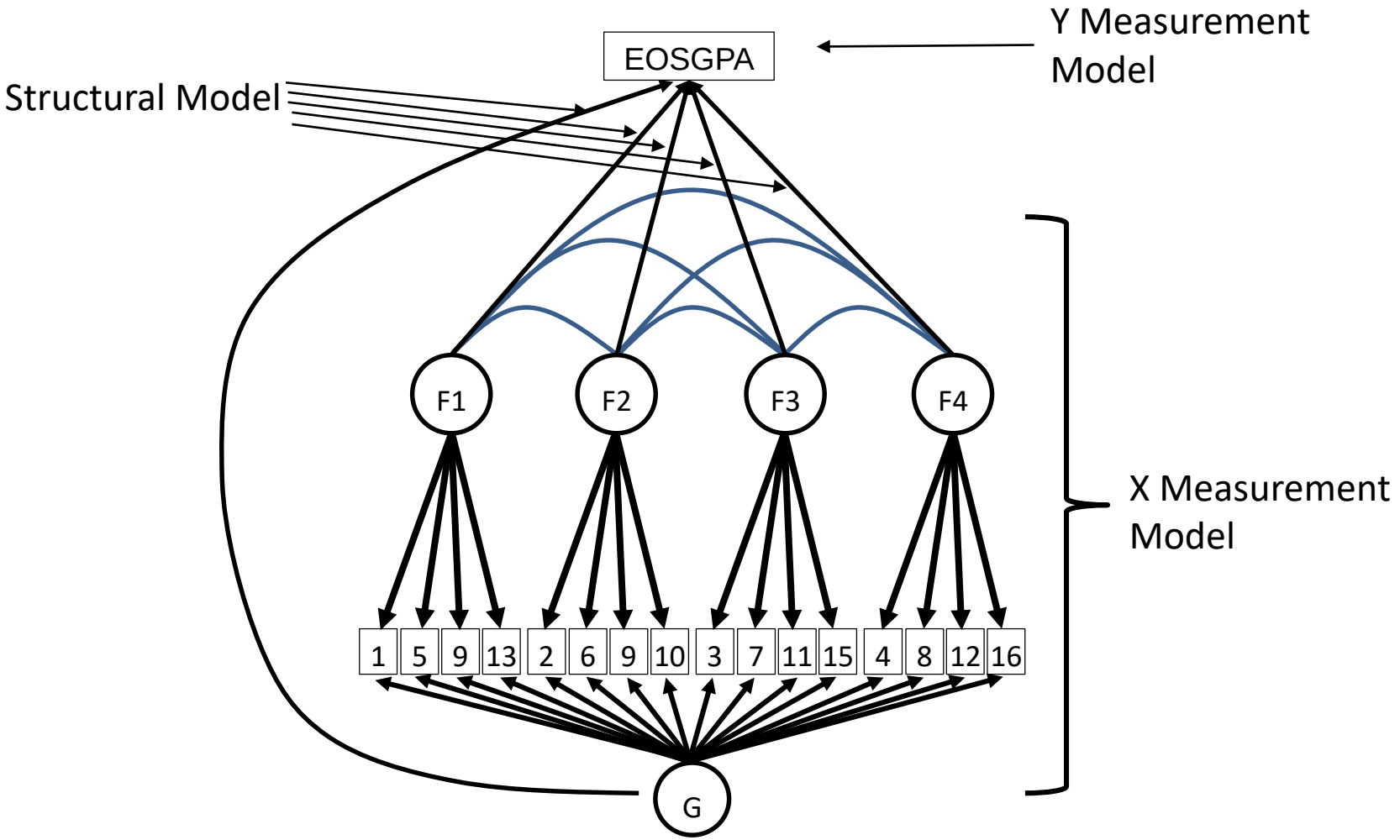
Note that there are no cross-loadings between the factor of the X measurement model and that of the Y measurement model.

SEM Example 1: Predicting GPA

- X Measurement Model – Our bifactor CFA of HEXACO Conscientiousness items.

- Y Measurement Model – Observed GPA at end of semester in which personality questionnaire was completed.

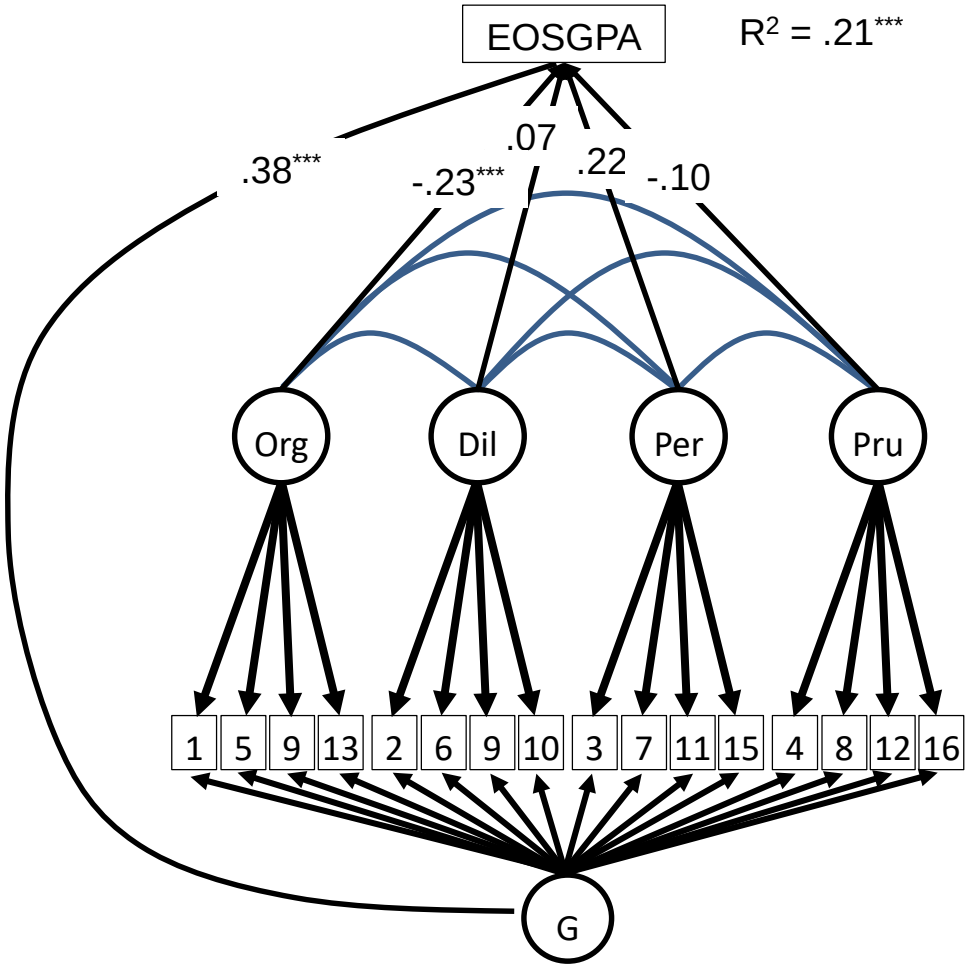
SEM Example 1 Path Diagram



SEM Example 1 Mplus Commands

```
title:  Spring 2014  HEXACO Data;
data:   FILE IS 'C:\Users\mdbid\Desktop\HEX GPA n1195 170409.csv';
variable:        names are
                 bidid
                 hh1 - hh16 hs1 - hs16
                 hx1 - hx16 ha1 - ha16
                 hc1 - hc16 ho1 - ho16 eosgpa ;
usevariables   are hc1 - hc16 eosgpa1;
analysis:     Type = general;
model:        f1 by hc1 hc5  hc9 hc13;
              f2 by hc2 hc6 hc10 hc14;
              f3 by hc3 hc7 hc11 hc15;
              f4 by hc4 hc8 hc12 hc16;
              g  by hc1 - hc16;
              g  with f1-f4@0;
             eosgpa on g f1 f2 f3 f4;    !THIS IS THE STRUCTURAL MODEL;
```

SEM Example 1 Results



Interpretational Confounding

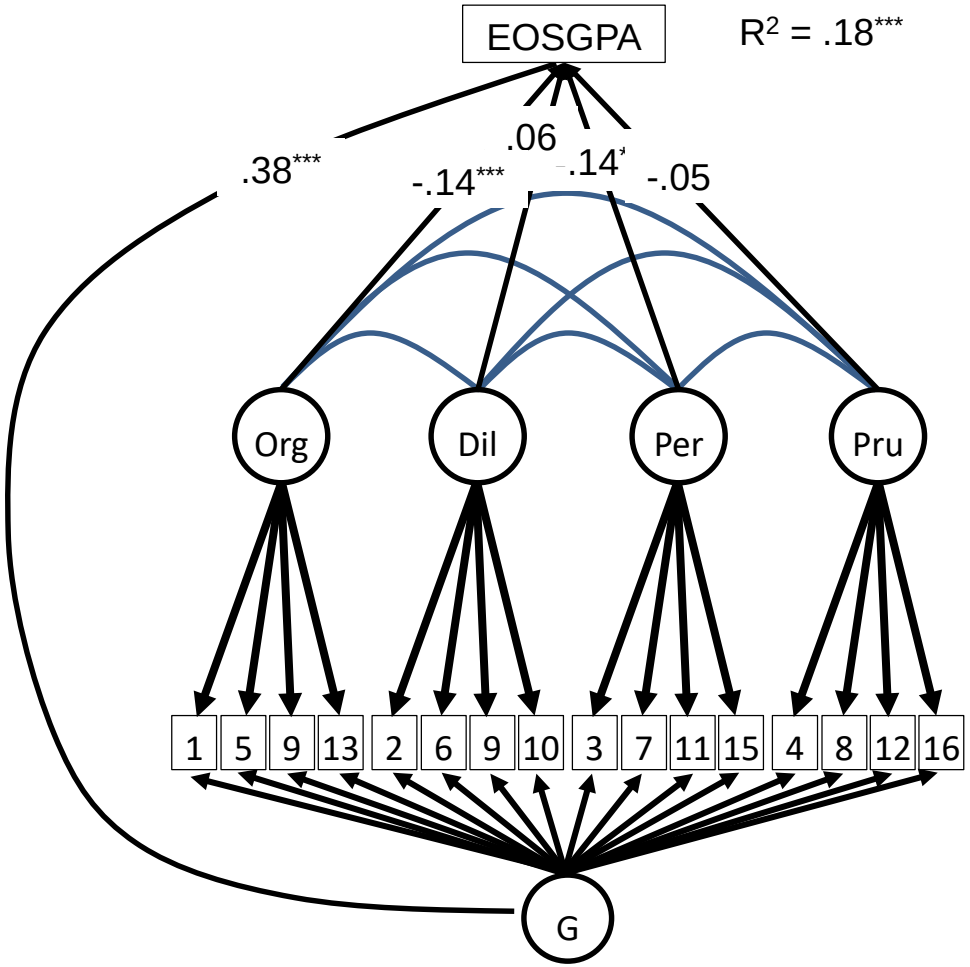
- It's possible that the estimates of the loadings defining a latent variable in either the X or Y measurement model may be affected by the presence of the structural relationship in the model.
- To check on that, I tested the same model, but used the loading estimates obtained from the measurement model without the structural model.

SEM Example 1 Conservative Mplus Commands

```
title:  Spring 2014  HEXACO Data;
data:   FILE IS 'C:\Users\mdbid\Desktop\HEX GPA n1195 170409.csv';
variable:        names are
                 bidid
                 hh1 - hh16 hs1 - hs16
                 hx1 - hx16 ha1 - ha16
                 hc1 - hc16 ho1 - ho16 eosgpa ;
usevariables   are hc1 - hc16 eosgpa1;
analysis:      Type = general;
model:         F1 BY  HC1@0.713  HC5 @0.239  HC9 @1.441  HC13@0.611;
               F2  BY HC2@0.393  HC6 @0.903  HC10@0.251  HC14@0.070;
               F3  BY HC3@0.230  HC7 @0.277  HC11@0.469  HC15@0.303;
               F4  BY HC4@0.851  HC8 @0.751  HC12@0.512  HC16@0.319;
                g  BY HC1@0.609  HC2 @0.648  HC3 @0.684  HC4 @0.500
                      HC5@0.967  HC6 @0.758  HC7 @0.747  HC8 @0.665
                      HC9@0.693  HC10@0.762  HC11@0.719  HC12@0.352
                      HC13@0.845 HC14@0.828  HC15@0.883  HC16@0.484;
                g with f1-f4@0;
               eosgpa on g f1 - f4;    !THIS IS THE STRUCTURAL MODEL;
```

From estimates obtained without the structural model.

SEM 1 Conservative Solution Results

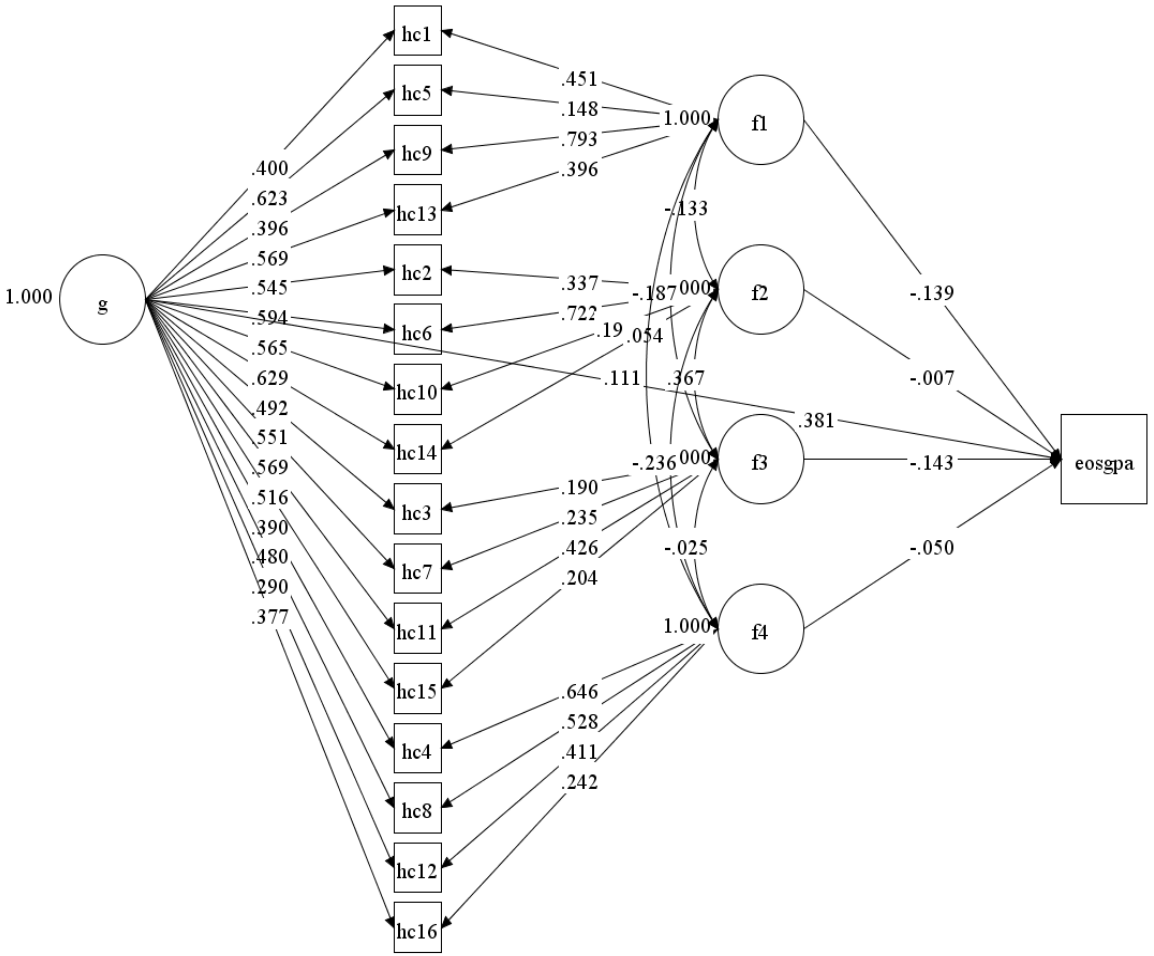


R^2 is slightly smaller.

Pattern of "significances" is the same.

See McAbee et al. (2014) for a similar analysis.

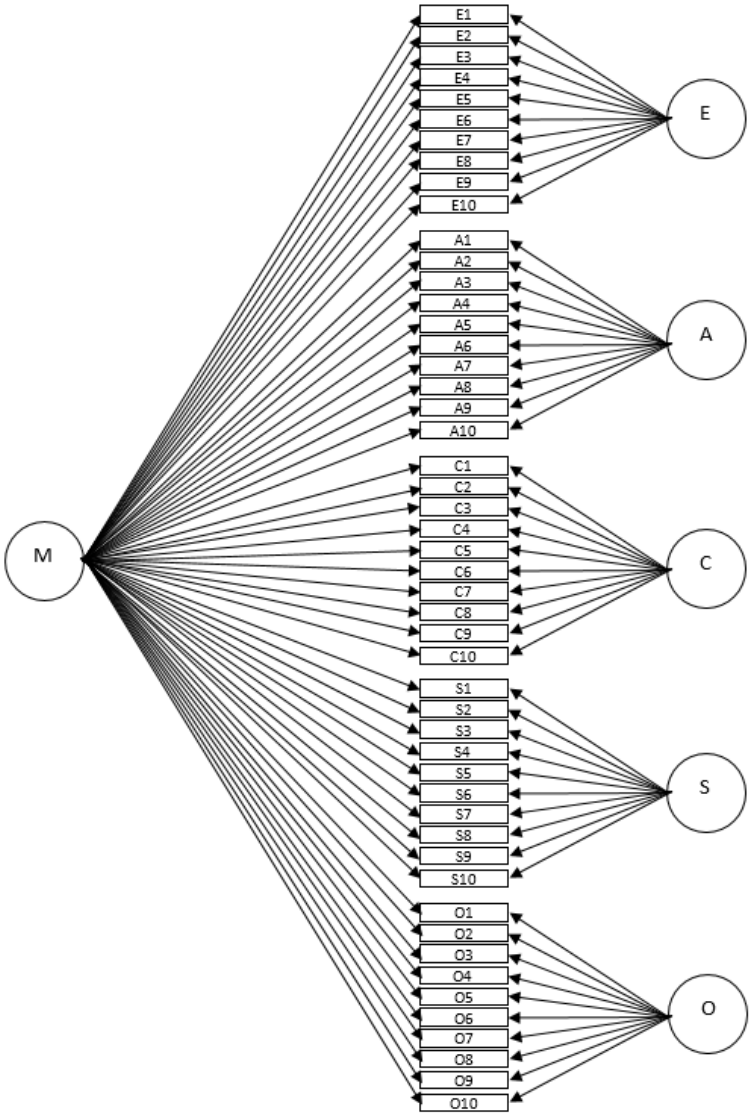
Mplus's Diagram of SEM 1



SEM Example 2: Predicting Supervisor Ratings

- Predicting Supervisor Ratings of performance of employees in a financial services company.
- Employees were rated on Service, Sales, and Collections. These were treated as indicators of a single Overall performance latent variable.
- Predictors were Big 5 Personality Factors and a general factor estimated from the Big 5 data

SEM Example 2
X Measurement Model

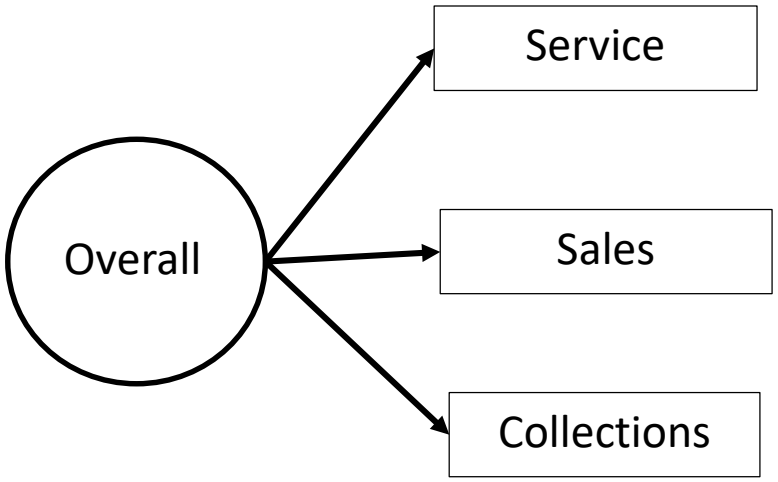


This is a bifactor CFA model of a Big Five Personality Questionnaire.

Factor correlations were estimated but not shown here to keep the figure simple.

We believe that M represents overall affect level or self regard of respondents.

SEM Example 2 – Y Measurement Model

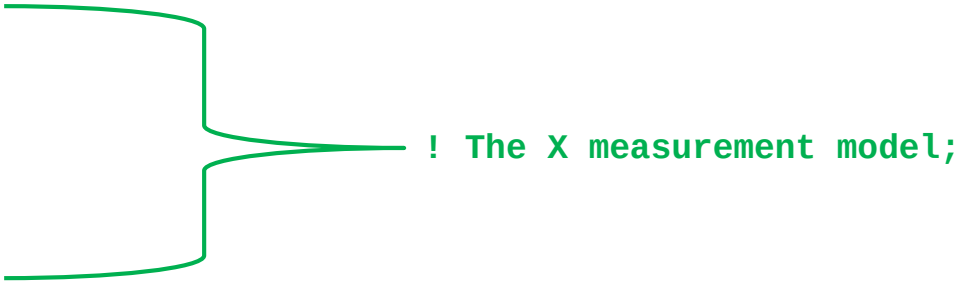


This is a 1-factor EFA/CFA model of the overall performance evaluations of employees at a walk-in financial services company.

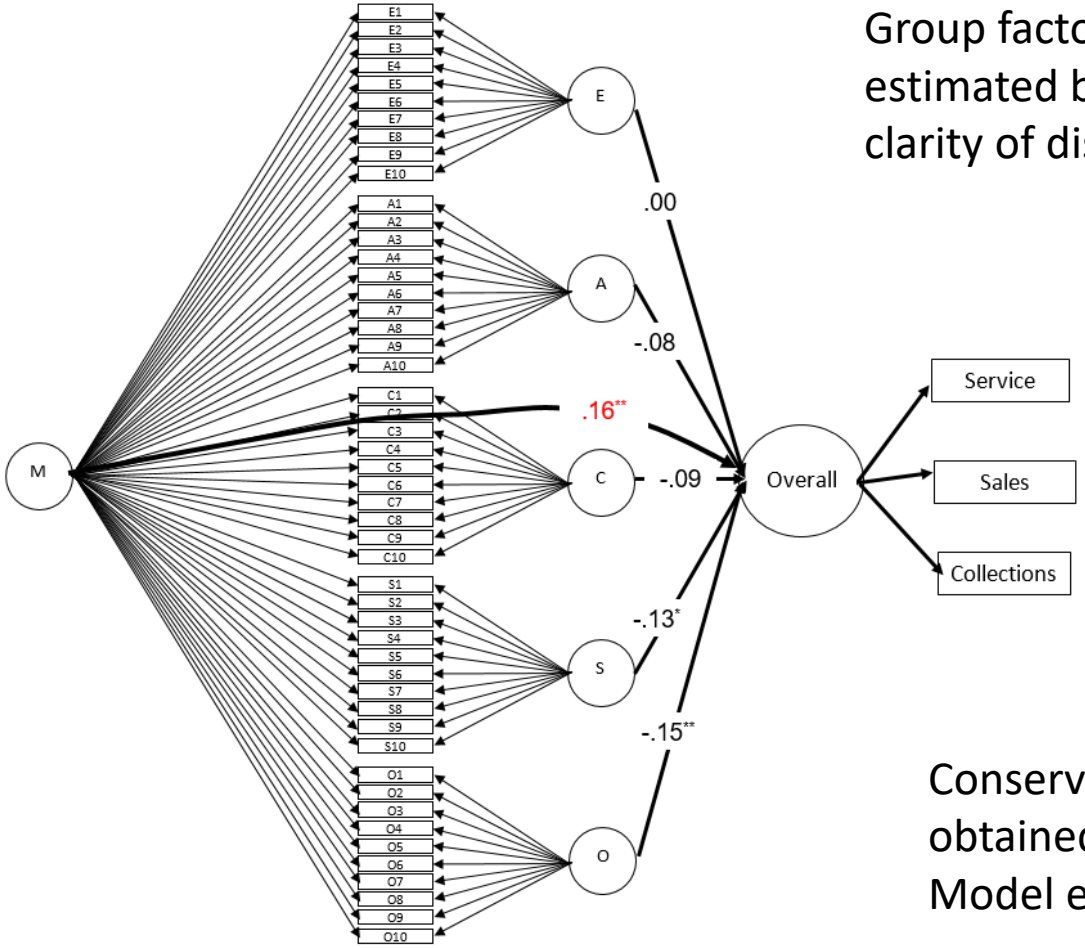
Supervisors rated each employee on Customer Service, Sales, and Collections.

Mplus SEM 2 Command File

```
title:  EAOM Mplus Workshop SEM example;

data:   FILE IS  'G:\MdbR\performance.csv';
        listwise=on;

variable:   names are e1-e10 a1-a10 c1-c10 s1-s10 o1-o10 custs sales colls;
usevariables are e1-o10 custs sales colls;

analysis:   type =  general ;

model:
            e by  e1-e10;
            a by  a1-a10;
            c by  c1-c10;
            s by  s1-s10;
            o by  o1-o10;
            m by  e1-o10;
            m with e-o@0;
            loverall by custs sales colls;     ! The Y measurement model;
            loverall on m e a c s o;           ! The structural model;
```



SEM Example 2 – SEM Results



Group factor correlations were estimated but are not shown for clarity of display.

Conservative results were obtained using Measurement Model estimates.

Summary of Main Points

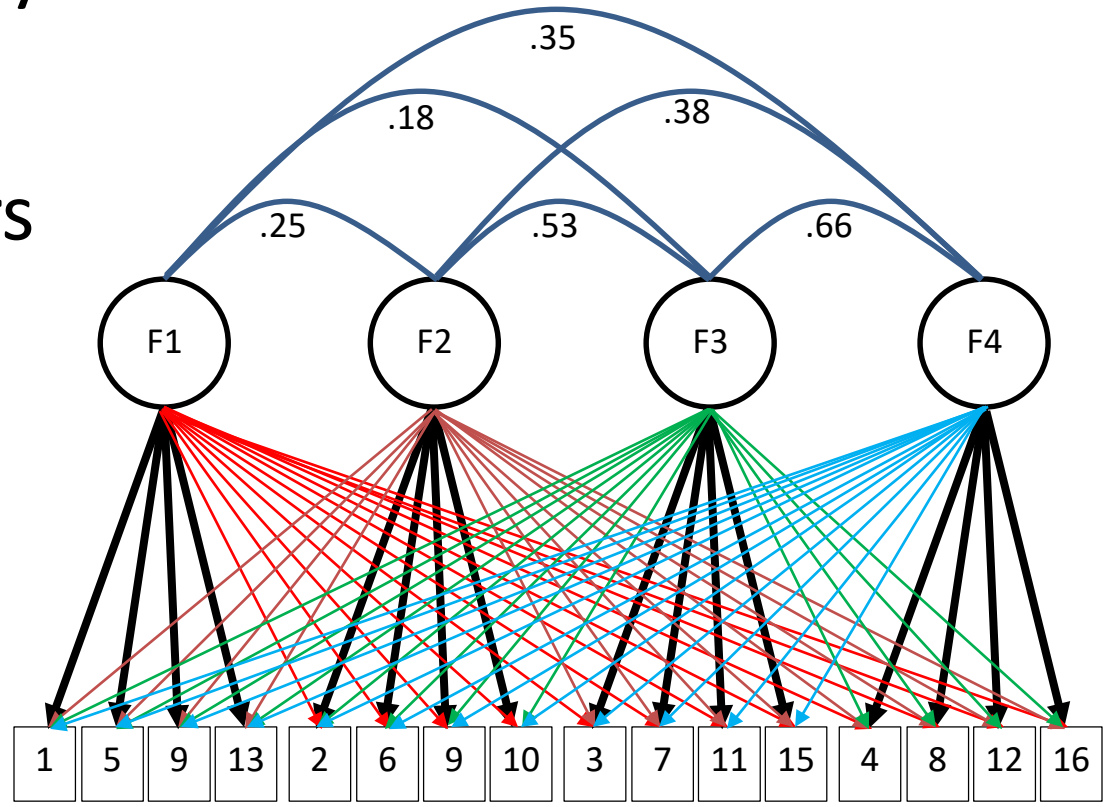
- Following are the main topics of this workshop

EFA: Exploratory Factor Analysis

- A factor analysis in which every factor influences every indicator
- Used to

  identify clusters

  of items

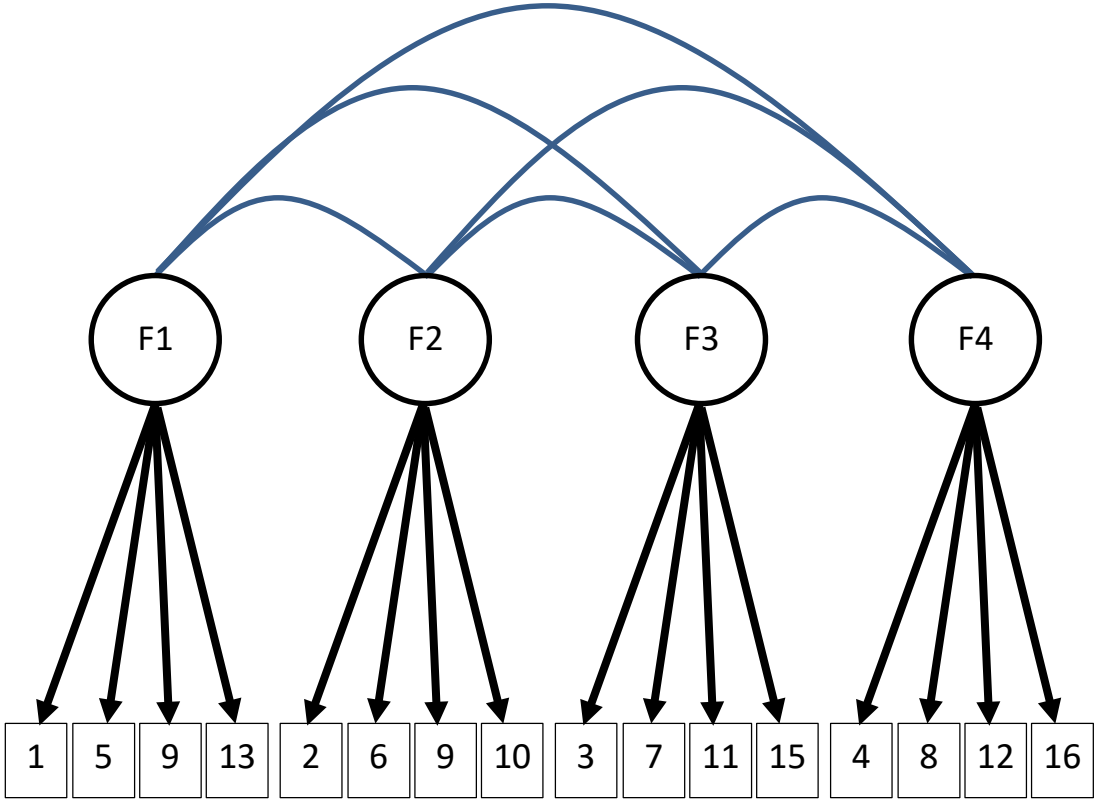


EFA Concepts

- Loadings
  - Principal loadings
  - Secondary loadings
- Factor correlations
- Path Diagram Symbols

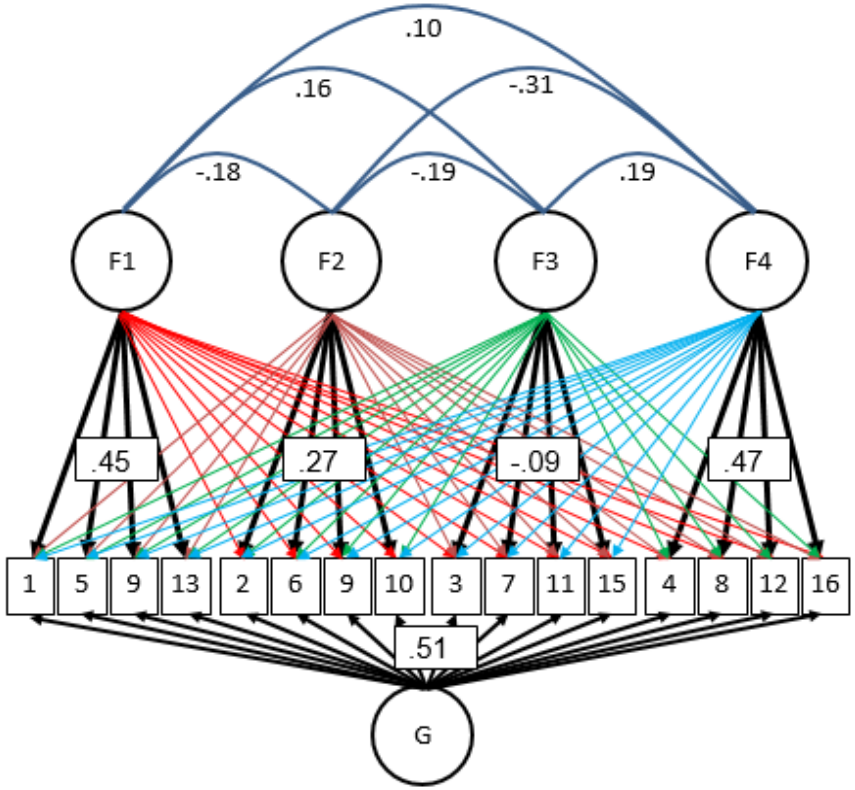
CFA

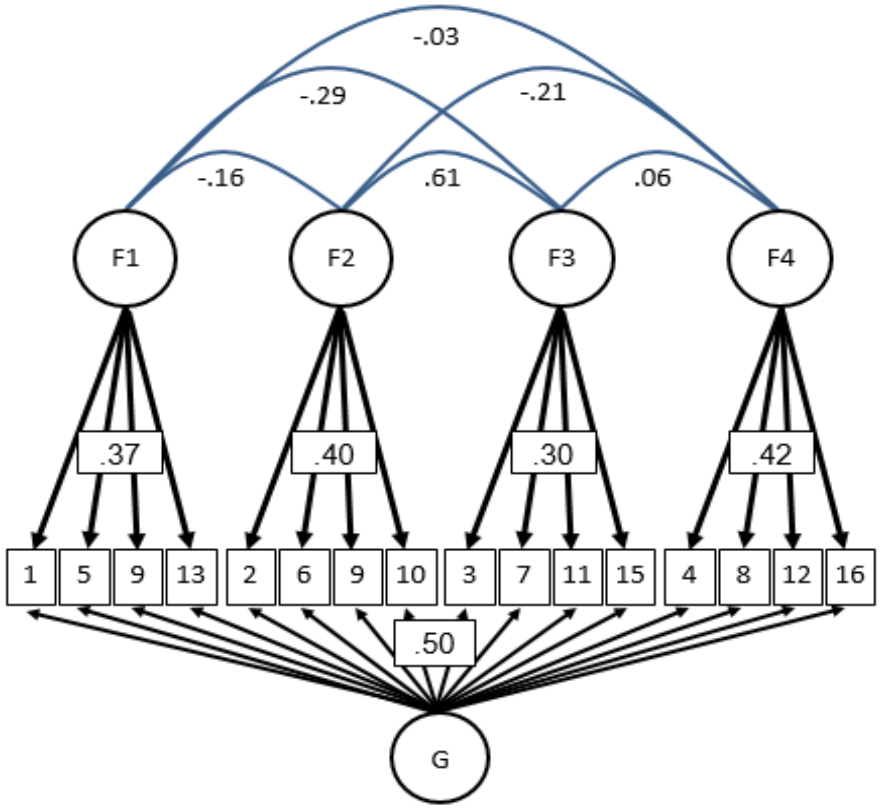
- A factor analysis in which some (OK,many) factor loadings are fixed, typically at zero.



Bifactor Solution

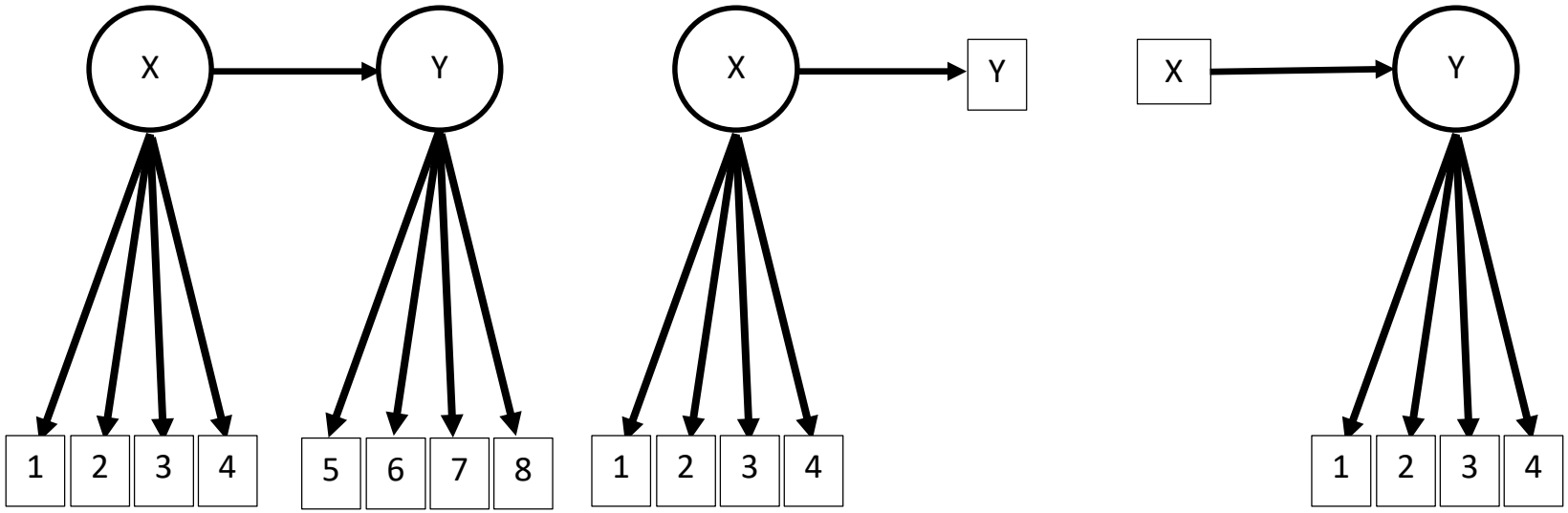
- A solution with a general factor that influences all indicators. Either EFA or CFA





Structural Equation Model

- A (usually) regression model involving latent variables as either X or Y or both.



Your turn

- Theory: Quality of relationships between supervisors and subordinates will affect productivity of employees.
- Quality here is based on Graen's LMX Scale
- Performance is based on 7 performance items
- Hypothesis: There will be a significant positive relationship between LMX and productivity.
- Your task: Fit a SEM model testing the relationship of LMX to productivity.
  - Due to the Mplus trial version's 8 variable limitation, please use only the first 5 indicators for LMX and 3 indicators for productivity variable.

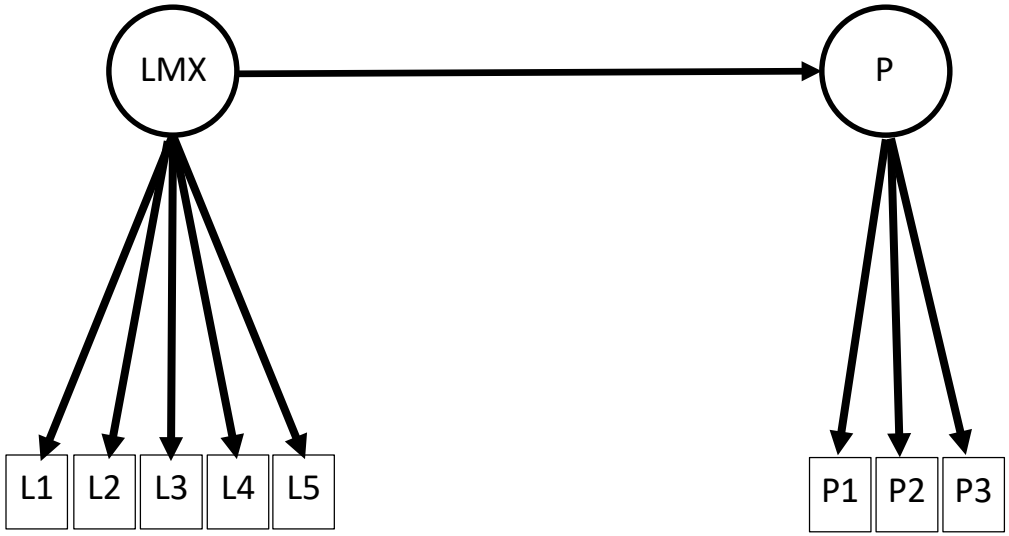
The data you'll analyze

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	3	2	3	3	3	2	2	5	5	5	5	5	4	4
2	4	3	4	4	1	2	2	5	5	5	4	4	5	5
3	3	2	4	3	3	3	2	3	3	3	3	3	3	5
4	3	3	3	3	3	2	3	5	5	4	5	4	5	4
5	2	2	4	4	3	3	2	3	3	3	3	4	4	5
6	2	2	2	2	2	2	3	4	5	4	5	5	4	4
7	3	3	4	2	4	4	4	4	4	4	4	4	4	4
8	2	2	2	2	1	2	2	3	3	3	3	3	3	5
9	3	3	3	1	2	4	2	3	4	3	3	4	5	4
10	3	3	4	1	1	3	3	5	5	5	5	5	5	5
11	4	3	4	4	4	4	4	4	5	5	5	5	5	5
12	3	3	2	4	2	2	2	2	3	2	2	2	2	3

- Data file is 'sem_eamnomissing.csv'
- Column A – G are LMX items.
- Columns H-N are Performance items

Path Diagram of In-class Model

- LMX is the Predictor (X).
- Performance is the criterion (Y)



References

Burt, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods and Research, 5*, 3-52.

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2*, 41-54.

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

McAbee, S. T., Oswald, F. L., & Connelly, B. S. (2014). Bifactor models of personality and college student performance: A broad versus narrow view. *European Journal of Personality, 28,* 604-619.

Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling, 23*, 116-139.

Muthén, L. K., & Muthén, B. O. (1998-2014). *Mplus User's Guid*e, Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling*. New York: Guilford.

Raykov, T., & Marcoulides, G. (2006). *A First Course in Structural Equation Modeling. 2nd Ed.* Erlbaum: Mahwah, NJ.

Williams, L.J., Ford, L. R., & Nguyen, N.T. (2002). Basic and Advanced Measurement Models for Confirmatory Factor Analysis. In S. Rogelberg (Ed.). *Handbook of Research Methods in Industrial and Organizational Psychology* (pp.366-389). Oxford: Blackwell.