STATISTICAL HORIZONS

Latent Growth Curve Modeling

Gregory Hancock, Ph.D.

Upcoming Seminar:

October 1-3, 2020, Remote Seminar

GREGORY R. HANCOCK
University of Maryland

LATENT GROWTH CURVE MODELING

TOPICS
Review of SEM and Software Basics
Mean Structure Models
Linear Model Foundations
Nonlinear Models
Other Cool Stuff
Sample Size Planning

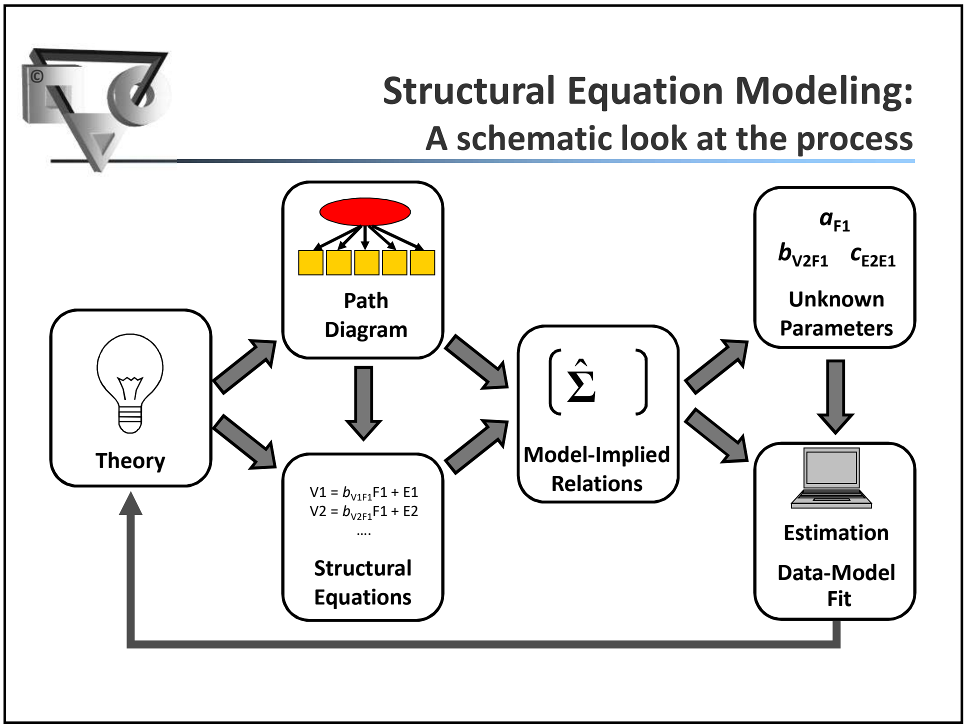
REVIEW OF SEM AND SOFTWARE BASICS

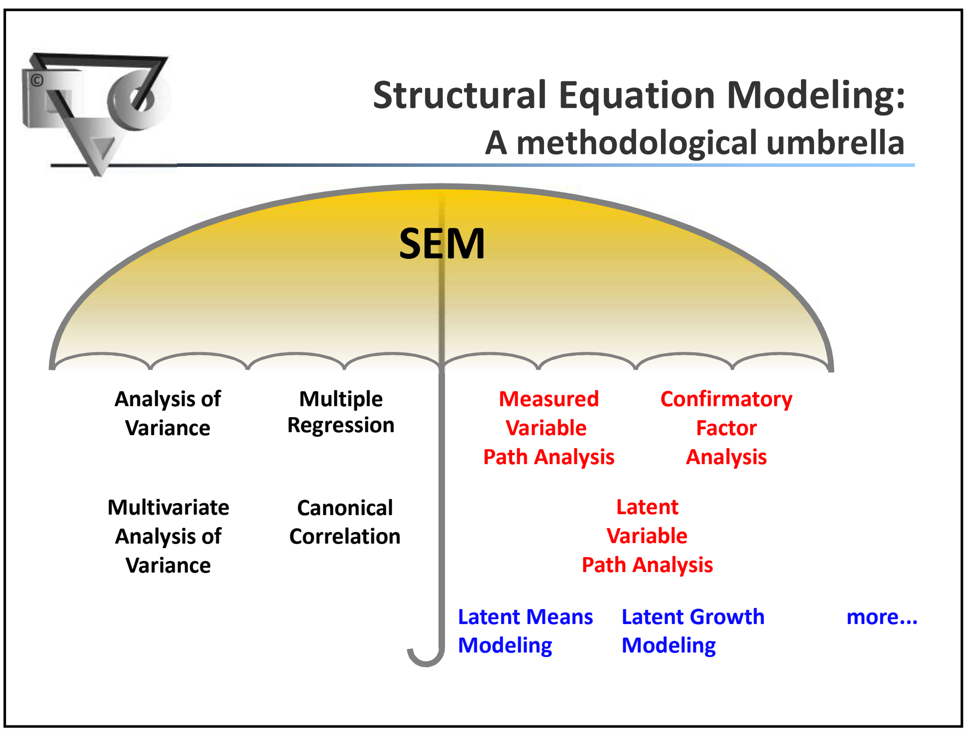
- orientation / notation
- data-model fit
- measured variable path models
- confirmatory factor models
- latent variable path models
- Mplus code / output
- resources

Structural Equation Modeling: A process to understand variables

Structural equation modeling is a process that allows for the assessment of (typically causal) theories involving measured and possibly latent variables to explain the characteristics of measured variables – variances, covariances, and sometimes means.

- Model Conceptualization
- Parameter Identification
- Parameter Estimation
- Data-Model Fit Assessment
- Possible Model Modification


Structural Equation Modeling:
A schematic look at the process
Theory
Path Diagram
V1 = b_{V1F1}F1 + E1
V2 = b_{V2F1}F1 + E2
....
Structural Equations
$\left(\hat{\Sigma}\right)$
Model-Implied Relations
a_{F1}
b_{V2F1} c_{E2E1}
Unknown Parameters
Estimation
Data-Model Fit


Structural Equation Modeling:
A methodological umbrella
SEM
Analysis of Variance
Multiple Regression
Measured Variable Path Analysis
Confirmatory Factor Analysis
Multivariate Analysis of Variance
Canonical Correlation
Latent Variable Path Analysis
Latent Means Modeling
Latent Growth Modeling
more...

Classic Notational Systems and Conventions

Bentler-Weeks (BW)		Jöreskog-Keesling-Wiley (JKW) ("LISREL" notation)	
V	observed Variable		x y
F	latent Factor	Exogenous: Endogenous:	ξ (ksi) η (eta)
E	Error (in Vs)		δ (delta) ε (epsilon)
D	Disturbance (in Fs)		ζ (zeta)

Parameter Notation System: Hancock-Mueller *a-b-c* parameter notation*

a

1

intercept; mean

b

direct structural effect

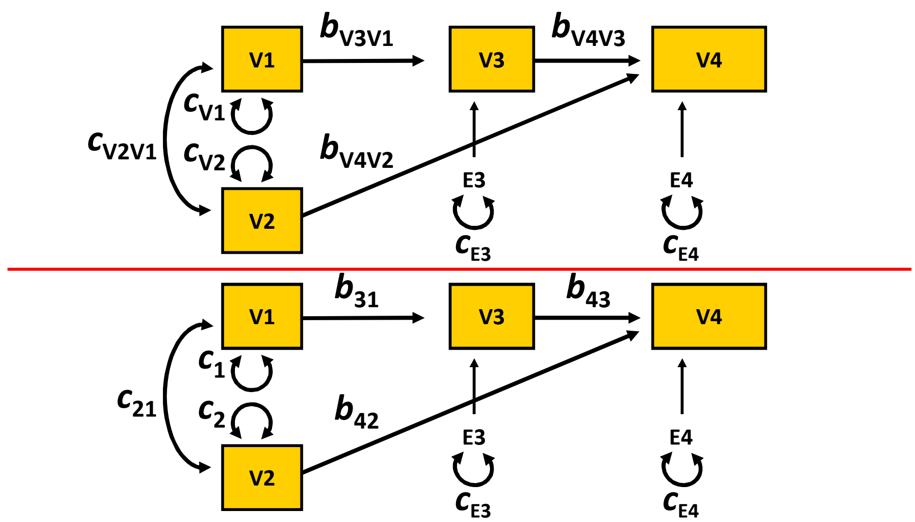
c

covariance between exogenous elements

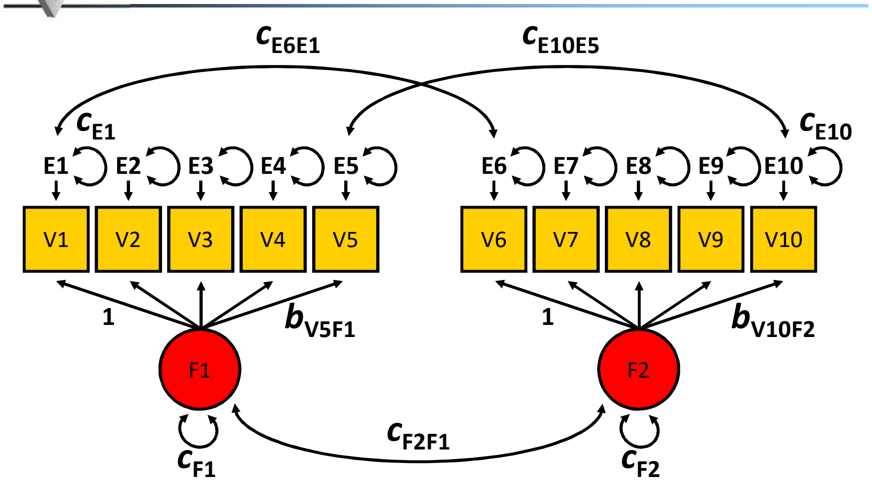
variance of exogenous element

* Hancock & Mueller, 2006/2013

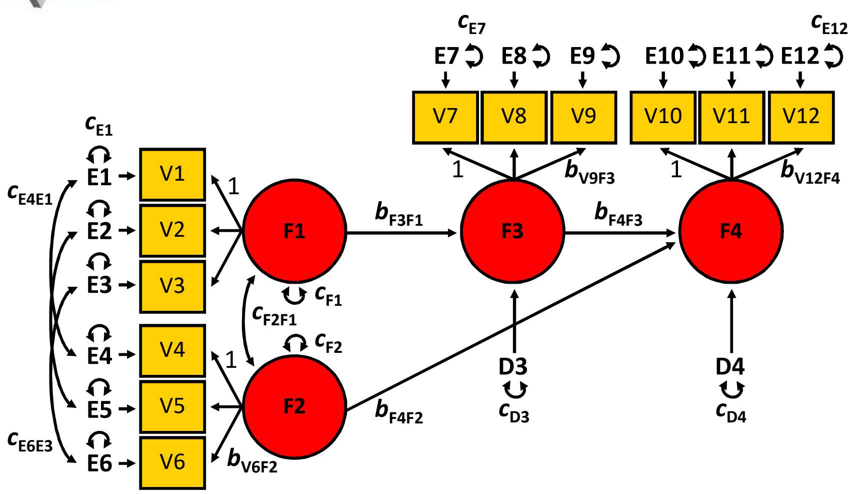
Measured Variable Path Analysis: *a-b-c* notation



Confirmatory Factor Analysis: *a-b-c* notation

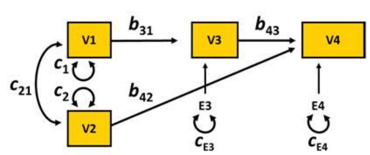


Latent Variable Path Analysis: *a-b-c* notation



Model-Implied Covariance Matrix

model



model-implied covariance matrix

$$\hat{\Sigma}=\begin{bmatrix} c_1 & & & \\ c_{21} & c_2 & & \\ b_{31}c_1 & c_{21}b_{31} & b_{31}^2c_1+c_{E3} & \\ b_{31}b_{43}c_1+c_{21}b_{42} & c_{21}b_{31}b_{43}+b_{42}c_2 & b_{43}(b_{31}^2c_1+c_{E3})+b_{31}c_{21}b_{42} & \begin{matrix} b_{43}^2(b_{31}^2c_1+c_{E3})+b_{42}^2c_2 \\ +2b_{42}c_{21}b_{31}b_{43}+c_{E4} \end{matrix} \end{bmatrix}$$

Observed vs. Model-Implied:
"Good" data-model fit?

$$\mathbf{S} = \begin{bmatrix} \quad \end{bmatrix} \qquad \hat{\Sigma} = \begin{bmatrix} \text{a function of all model parameters} \end{bmatrix}$$

how the data actually behave in the sample

how the model implies the data *should* behave in the population

Good data-model fit means that we can find estimates of the parameters in the model (through, for example, maximum likelihood estimation), to get a model-implied covariance matrix that is reasonably close to the observed covariance matrix.

But how do we measure "close"?

Data-Model Fit Assessment:
A selection of indices

Absolute (observed vs. model-implied var/cov matrix)	**Parsimonious** (adjust for model complexity)	**Incremental** (target vs. baseline model)
Model χ^2 statistic	**Akaike Information Criterion (AIC)**	**Comparative Fit Index (CFI)**
Standardized Root Mean Squared Residual (SRMR)	**Root Mean Squared Error of Approximation (RMSEA)**	**Normed Fit Index (NFI)**
Goodness-of-Fit Index (GFI)	**Adjusted Goodness-of-Fit Index (AGFI)**	**Nonnormed Fit Index (NNFI; also known as Tucker-Lewis Index)**

Data-Model Fit Assessment: Absolute

- $$\text{SRMR} = \sqrt{\frac{\sum_{1}^{u}(\text{model-based standardized residual})^2}{u}}$$

  where $u = p(p+1)/2$ is the number of unique variances/covariances among the p variables in the model

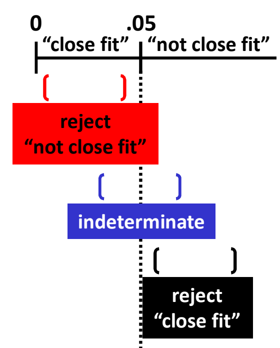
- Smaller values indicate better fit.

Data-Model Fit Assessment: Parsimonious

- $$\text{RMSEA} = \sqrt{\max\left[\frac{(\chi^2/df)-1}{n-1}, 0\right]}$$

- Smaller values indicate better fit.

- Comes with a confidence interval, typically a 90% CI.



Data-Model Fit Assessment: Incremental

- $$\mathrm{CFI} = 1 - \frac{\max\left[\left(\chi^2_{\text{model}} - df_{\text{model}}\right), 0\right]}{\max\left[\left(\chi^2_{\text{null}} - df_{\text{null}}\right), \left(\chi^2_{\text{model}} - df_{\text{model}}\right), 0\right]}$$

$$\mathrm{CFI} = 1 - \frac{\chi^2_{\text{model}} - df_{\text{model}}}{\chi^2_{\text{null}} - df_{\text{null}}}$$ **almost always**

Data-Model Fit Assessment: Hu & Bentler (1999)

Absolute	Parsimonious	Incremental
SRMR≤.08	RMSEA ≤.06	CFI ≥.95

The values are not set in stone. They were derived based on a broad, but not unlimited, set of models.

They should be treated as guidelines, **not laws**.

Data-Model Fit Assessment:
Interpretation

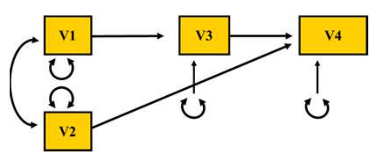
- Poor data-model fit?

  Reject the hypothesized model. Entertain modifications *only* if they make theoretical and statistical sense.

- Satisfactory data-model fit?

  Tentatively retain the proposed model as *one* viable representation of the true relations underlying the data.

Measured Variable Path Analysis:
Mplus syntax




```
DATA:
  FILE IS mvpa_data.csv;
VARIABLE:
  NAMES ARE V1 V2 V3 V4;
ANALYSIS:                  } default
  ESTIMATOR IS ML;
MODEL:
  V3 ON V1;
  V4 ON V2 V3;
  V1; V2; V3; V4;          } default
  V1 WITH V2;
OUTPUT:
  SAMPSTAT STANDARDIZED;
```

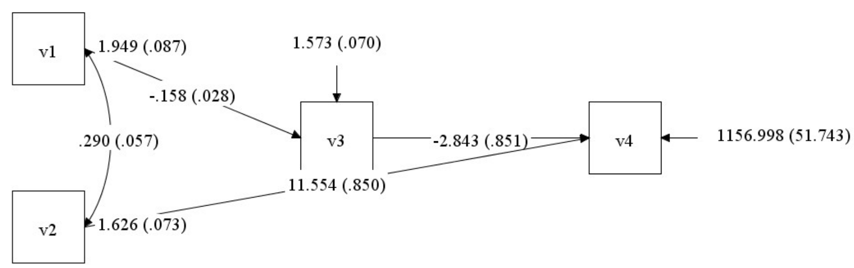
Measured Variable Path Analysis: Mplus output

```
MODEL FIT INFORMATION

Chi-Square Test of Model Fit
          Value                          360.423
          Degrees of Freedom                   2
          P-Value                         0.0000

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                         0.423
          90 Percent C.I.                  0.387  0.461
          Probability RMSEA <= .05         0.000

CFI/TLI
          CFI                              0.357
          TLI                             -0.607

SRMR (Standardized Root Mean Square Residual)
          Value                            0.137
```

Measured Variable Path Analysis: Mplus output

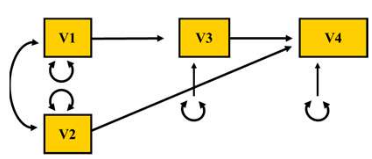
```
                                                  Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 V3       ON
    V1                -0.158      0.030     -5.282      0.000

 V4       ON
    V2                11.554      0.872     13.248      0.000
    V3                -2.843      0.830     -3.424      0.001

 V1       WITH
    V2                 0.290      0.060      4.864      0.000

 Variances
    V1                 1.949      0.072     26.991      0.000
    V2                 1.626      0.061     26.664      0.000

 Residual Variances
    V3                 1.573      0.057     27.534      0.000
    V4              1156.998     51.711     22.374      0.000
```

Measured Variable Path Analysis:
Mplus output



Measured Variable Path Analysis:
Mplus syntax for indirect effects

```
DATA:
  FILE IS mvpa_data.csv;
VARIABLE:
  NAMES ARE V1 V2 V3 V4;
ANALYSIS:
  BOOTSTRAP = 1000;
MODEL:
  V3 ON V1;
  V4 ON V2 V3;
  V1; V2; V3; V4;
  V1 WITH V2;
MODEL INDIRECT:
  V4 IND V3 V1;
OUTPUT:
  SAMPSTAT STANDARDIZED CINTERVAL(BCBOOTSTRAP);
```



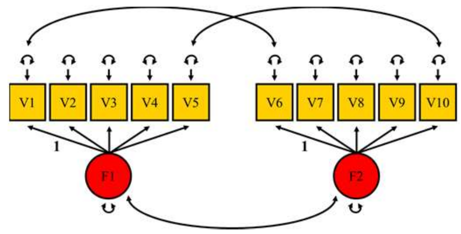
Measured Variable Path Analysis: Mplus output

```
CONFIDENCE INTERVALS OF TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

                  Lower .5%  Lower 2.5%    Lower 5%    Estimate    Upper 5%  Upper 2.5%   Upper .5%

Effects from V1 to V4

  Sum of indirect   -0.038       0.079       0.138       0.448       0.759       0.818       0.935

  Specific indirect

    V4
    V3
    V1              -0.038       0.079       0.138       0.448       0.759       0.818       0.935
```

Measured Variable Path Analysis: Mplus output

```
CONFIDENCE INTERVALS OF TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

                  Lower .5%  Lower 2.5%    Lower 5%    Estimate    Upper 5%  Upper 2.5%   Upper .5%

Effects from V1 to V4

  Indirect           0.071       0.151       0.194       0.448       0.809       0.906       1.014



CONFIDENCE INTERVALS OF STANDARDIZED TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS


STDYX Standardization

                  Lower .5%  Lower 2.5%    Lower 5%    Estimate    Upper 5%  Upper 2.5%   Upper .5%

Effects from V1 to V4

  Indirect           0.003       0.005       0.007       0.017       0.030       0.033       0.038
```

Confirmatory Factor Analysis:
Mplus syntax




```
DATA:
  FILE IS cfa_data.csv;
VARIABLE:
  NAMES ARE V1-V10;
MODEL:
  F1 BY V1-V5;      first loading
  F2 BY V6-V10;     set to 1 by default
  V1 WITH V6;
  V5 WITH V10;
  V1-V10;
  F1-F2;           default
  F1 WITH F2;
OUTPUT:
  SAMPSTAT STANDARDIZED;
```

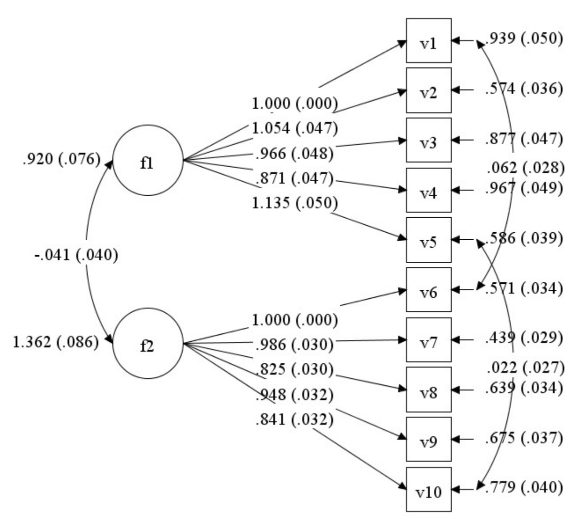
Confirmatory Factor Analysis:
Mplus output

```
MODEL FIT INFORMATION

Chi-Square Test of Model Fit
          Value                              32.544
          Degrees of Freedom                     32
          P-Value                            0.4400

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                            0.004
          90 Percent C.I.                     0.000  0.024
          Probability RMSEA <= .05            1.000

CFI/TLI
          CFI                                 1.000
          TLI                                 1.000

SRMR (Standardized Root Mean Square Residual)
          Value                               0.018
```

Confirmatory Factor Analysis: Mplus output

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

 F1       BY
    V1              1.000      0.000    999.000    999.000
    V2              1.054      0.047     22.369      0.000
    V3              0.966      0.048     20.084      0.000
    V4              0.871      0.047     18.500      0.000
    V5              1.135      0.050     22.831      0.000

 F2       BY
    V6              1.000      0.000    999.000    999.000
    V7              0.986      0.030     33.098      0.000
    V8              0.825      0.030     27.828      0.000
    V9              0.948      0.032     29.610      0.000
    V10             0.841      0.032     26.627      0.000

 F1       WITH
    F2             -0.041      0.040     -1.017      0.309
```

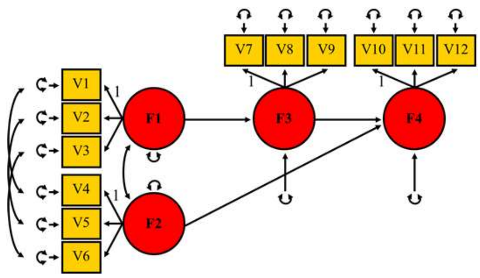
Confirmatory Factor Analysis: Mplus output



Latent Variable Path Analysis:
Mplus syntax

```
DATA:
  FILE IS lvpa_data.txt;
VARIABLE:
  NAMES ARE V1-V12;
MODEL:
  F1 BY V1-V3;
  F2 BY V4-V6;
  F3 BY V7-V9;
  F4 BY V10-V12;
  V1-V3 PWITH V4-V6;
  F3 ON F1;
  F4 ON F2 F3;
  V1-V12; F1-F4;
  F1 WITH F2;
OUTPUT:
  SAMPSTAT STANDARDIZED;
```

default



Latent Variable Path Analysis:
Mplus output

```
MODEL FIT INFORMATION

Chi-Square Test of Model Fit
          Value                            129.639
          Degrees of Freedom                    47
          P-Value                           0.0000

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.042
          90 Percent C.I.                    0.033  0.051
          Probability RMSEA <= .05           0.936

CFI/TLI
          CFI                                0.983
          TLI                                0.977

SRMR (Standardized Root Mean Square Residual)
          Value                              0.154
```

Latent Variable Path Analysis:
Mplus output

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
F3	ON				
F1		0.300	0.049	6.174	0.000
F4	ON				
F2		19.710	0.875	22.525	0.000
F3		15.508	1.197	12.960	0.000
F1	WITH				
F2		-0.184	0.041	-4.487	0.000
V1	WITH				
V4		-0.035	0.033	-1.065	0.287
V2	WITH				
V5		-0.015	0.029	-0.497	0.619
V3	WITH				
V6		0.012	0.032	0.385	0.700

Latent Variable Path Analysis:
Mplus output