

---

# **Multifactor Confirmatory Factor Analysis**

Latent Trait Measurement and  
Structural Equation Models

Lecture #9

March 13, 2013

# Today's Class

---

- Confirmatory Factor Analysis with more than one factor:
  - Logic of multiple factors
    - ◆ By design
    - ◆ By model modification
  - Model comparison: nested and non-nested models
  - Cross-loaded items
  - Method factors
  - Additional estimation warnings and concerns
  - Structural models for factor covariances
    - ◆ Factor analysis of the factors

# Data for Today's Class

---

- Data were collected from two sources:
  - 144 “experienced” gamblers
    - ♦ Many from an actual casino
  - 1192 college students from a “rectangular” midwestern state
    - ♦ Many never gambled before
- Today, we will combine both samples and treat them as homogenous – one sample of 1346 subjects
  - Later we will test this assumption – measurement invariance (called differential item functioning in item response theory literature)
- We will build a multi-factor scale of gambling tendencies using 41 items of the GRI
  - Focused on long-term gambling tendencies

# Pathological Gambling: DSM Definition

---

- To be diagnosed as a pathological gambler, an individual must meet 5 of 10 defined criteria:

1. Is preoccupied with gambling
2. Needs to gamble with increasing amounts of money in order to achieve the desired excitement
3. Has repeated unsuccessful efforts to control, cut back, or stop gambling
4. Is restless or irritable when attempting to cut down or stop gambling
5. Gambles as a way of escaping from problems or relieving a dysphoric mood
6. After losing money gambling, often returns another day to get even
7. Lies to family members, therapist, or others to conceal the extent of involvement with gambling
8. Has committed illegal acts such as forgery, fraud, theft, or embezzlement to finance gambling
9. Has jeopardized or lost a significant relationship, job, educational, or career opportunity because of gambling
10. Relies on others to provide money to relieve a desperate financial situation caused by gambling

## Final 12 Items on the Scale

---

Item	Criterion	Question
GRI1	3	I would like to cut back on my gambling.
GRI3	6	If I lost a lot of money gambling one day, I would be more likely to want to play again the following day.
GRI5	2	I find it necessary to gamble with larger amounts of money (than when I first gambled) for gambling to be exciting.
GRI6	8	I have gone to great lengths to obtain money for gambling.
GRI9	4	I feel restless when I try to cut down or stop gambling.
GRI10	1	It bothers me when I have no money to gamble.
GRI11	5	I gamble to take my mind off my worries.
GRI13	3	I find it difficult to stop gambling.
GRI14	2	I am drawn more by the thrill of gambling than by the money I could win.
GRI15	7	I am private about my gambling experiences.
GRI21	1	It is hard to get my mind off gambling.
GRI23	5	I gamble to improve my mood.

# GRI 12 Item Analysis

- The 12 item analysis gave this model fit information:

## Chi-Square Test of Model Fit

Value	185.178*
Degrees of Freedom	54
P-Value	0.0000
Scaling Correction Factor for MLR	1.6034

The model  $\chi^2$  indicated the model did not fit better than the saturated model – but this statistic can be overly sensitive

## RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.043	
90 Percent C.I.	0.036	0.050
Probability RMSEA <= .05	0.949	

The model RMSEA indicated good model fit (want this to be < .05)

## CFI/TLI

CFI	0.952
TLI	0.941

The model CFI and TLI indicated the model fit well (want these to be > .95)

## SRMR (Standardized Root Mean Square Residual)

Value	0.032
-------	-------

The SRMR indicated the fit well (want this to be < .08)

- Additionally, only 4 normalized residuals were significant
  - Item 15 with item 1
  - Item 14 with item 3
  - Item 21 with item 3
  - Item 15 with item 6

---

# MULTIFACTOR CFA MODELS

# Multifactor CFA Models

---

- Multifactor CFA models are measurement models that measure more than one latent trait simultaneously
- The multiple factors represent theoretical constructs that are best when they are defined *a priori*
  - We will look at the entire 41 item GRI analysis with the intent of measuring 10 factors – one for each DSM criterion
  - But first...
- Multiple factors can also come from modification of models with single factors when model misfit is identified
  - We will begin with this to demonstrate with a smaller data set



# GRI 12 Item Scale: Revisiting Model Fit

---

- Although we deemed the 12-item GRI scale to fit a 1-factor model, there were a couple normalized residual covariances that were larger than 2:
  - GRI1 and GRI15: 2.434
  - GRI14 and GRI11: 3.643
  - GRI23 and GRI11: 2.092
- Positive normalized residual covariances: items are more related than your model predicts them to be
  - Something other than the factor created the relationship

# 1-Factor Model Covariances with Estimates

---

- For a pair of items,  $i$  and  $j$  the model-implied covariance in a 1-factor model:  $\lambda_{i1}\lambda_{j1}\sigma_{F_1}^2$
- From our model:
  - $\lambda_{14,1} = 1.023$
  - $\lambda_{23,1} = 1.086$
  - $\lambda_{11,1} = 1.012$
  - $\sigma_{F_1}^2 = 0.331$
- The model-implied covariance are:
  - $\sigma_{14,11} = 1.023 * 1.012 * 0.331 = 0.343$
  - $\sigma_{23,11} = 1.086 * 1.012 * 0.331 = 0.364$
- The actual (H1 model) covariances for these items were:
  - $\sigma_{14,11} = 0.539$
  - $\sigma_{23,11} = 0.458$

# Ways of Fixing The Model #1:

## Adding Parameters – Increasing Complexity

---

- A common source of misfit is due to items that have a significant residual covariance: items are still correlated after accounting for the common factor
- Solutions that increase model complexity:
  - **Add additional factors** (recommended solution)
    - ♦ Additional factors can be hard to specify
    - ♦ Model implied covariance adds terms from the other factor(s)
      - Items with “simple structure” (measuring only one factor) have additive terms
      - Items with cross-loadings are more complex
  - **Add a residual covariance between items** (dangerous solution)
    - ♦ Use modification indices to determine which to add
    - ♦ Error covariances are unaccounted for multi-dimensionality
      - This means you have measured your factor **and** something else that those items have in common (e.g. stem, valence, specific content, additional factors)
    - ♦ The model implied covariance then becomes  $\lambda_{i1}\lambda_{j1}\sigma_{F_1}^2 + \sigma_{ij}$ 
      - Virtually no residual as the additive term can mold to whatever it needs to be

# Adding an Additional Factor

---

- A justifiable reason must exist to add another factor
  - Good: Items appear to measure some different construct
  - Bad: Model did not fit well
- We are adding another factor based on the bad reason
  - Used to demonstrate how multifactor CFA works
- In practice we would first inspect the misfitting items and see if there are any similarities:
  - **11. I gamble to take my mind off my worries.**
  - **14. I am drawn more by the thrill of gambling than by the money I could win.**
  - **23. I gamble to improve my mood.**

## 2-Factor 12-Item GRI Model: Theoretical Implications

---

The CFA model for the 12 GRI items:

$$\begin{aligned}Y_{S,1} &= \mu_{I_1} + \lambda_{1,1}F_{S1} + 0 * F_{S2} + e_{S,1} \\Y_{S,3} &= \mu_{I_3} + \lambda_{3,1}F_{S1} + 0 * F_{S2} + e_{S,3} \\Y_{S,5} &= \mu_{I_5} + \lambda_{5,1}F_{S1} + 0 * F_{S2} + e_{S,5} \\Y_{S,6} &= \mu_{I_6} + \lambda_{6,1}F_{S1} + 0 * F_{S2} + e_{S,6} \\Y_{S,9} &= \mu_{I_9} + \lambda_{9,1}F_{S1} + 0 * F_{S2} + e_{S,9} \\Y_{S,10} &= \mu_{I_{10}} + \lambda_{10,1}F_{S1} + 0 * F_{S2} + e_{S,10} \\Y_{S,11} &= \mu_{I_{11}} + 0 * F_{S1} + \lambda_{11,2}F_{S2} + e_{S,11} \\Y_{S,13} &= \mu_{I_{13}} + \lambda_{13,1}F_{S1} + 0 * F_{S2} + e_{S,13} \\Y_{S,14} &= \mu_{I_{14}} + 0 * F_{S1} + \lambda_{14,2}F_{S2} + e_{S,14} \\Y_{S,15} &= \mu_{I_{15}} + \lambda_{15,1}F_{S1} + 0 * F_{S2} + e_{S,15} \\Y_{S,21} &= \mu_{I_{21}} + \lambda_{21,1}F_{S1} + 0 * F_{S2} + e_{S,21} \\Y_{S,23} &= \mu_{I_{23}} + 0 * F_{S1} + \lambda_{23,2}F_{S2} + e_{S,23}\end{aligned}$$

# Model Details

---

- As with the 1-factor model:
  - $Y_{si}$  - response of subject  $s$  on item  $i$
  - $\mu_{I_i}$  - intercept of item  $i$  (listed as a mean as this is typically what it becomes)
  - $\lambda_{if}$  - factor loading of item  $i$  on factor  $f$  ( $f = 1$  or  $2$ )
  - $F_{sf}$  - latent “factor score” for subject  $s$  (same for all items) for factor  $f$
  - $e_{si}$  - regression-like residual for subject  $s$  on item  $i$ 
    - ♦ We assume  $e_{si} \sim N(0, \psi_i^2)$ ;  $\psi_i^2$  is called the **unique variance** of item  $i$
    - ♦ We also assume  $e_{si}$  and  $F_{sf}$  are independent for all factors

## Model Assumptions: Latent Factors (Random Effects)

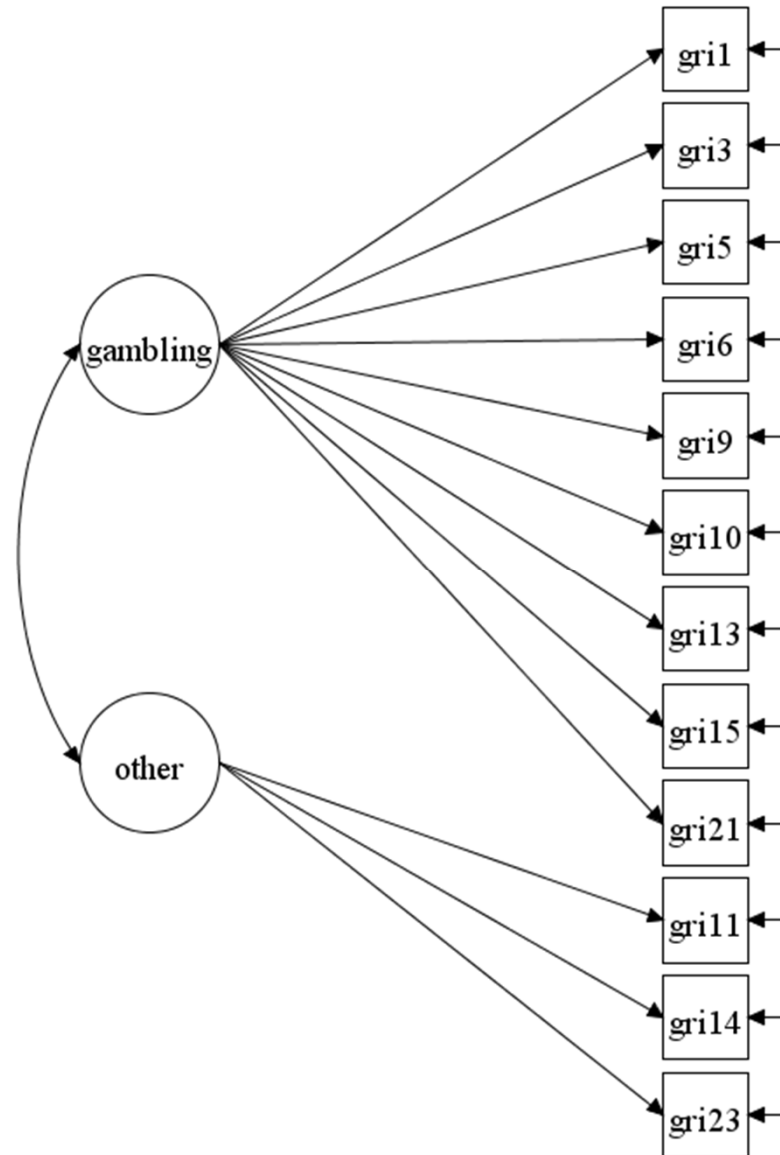
---

$$\begin{bmatrix} F_{S1} \\ F_{S2} \end{bmatrix} \sim N_2 \left( \boldsymbol{\mu}_F = \begin{bmatrix} \mu_{F_1} \\ \mu_{F_2} \end{bmatrix}, \boldsymbol{\Phi} = \begin{bmatrix} \sigma_{F_1}^2 & \sigma_{F_1, F_2} \\ \sigma_{F_1, F_2} & \sigma_{F_2}^2 \end{bmatrix} \right)$$

- Depending on how you identify the model, these terms are either estimated or fixed, most common choices are:
- Zero factor mean, estimated factor variance (marker items – what our examples will use for the most part):
  - $\boldsymbol{\mu}_F = \mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \boldsymbol{\Phi} = \begin{bmatrix} \sigma_{F_1}^2 & \sigma_{F_1, F_2} \\ \sigma_{F_1, F_2} & \sigma_{F_2}^2 \end{bmatrix}$
  - One loading for each factor is fixed (i.e.,  $\lambda_{1,1} = 1$  and  $\lambda_{11,2} = 1$ )
- Zero factor mean, standardized factors:
  - $\boldsymbol{\mu}_F = \mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \boldsymbol{\Phi} = \begin{bmatrix} 1 & \sigma_{F_1, F_2} \\ \sigma_{F_1, F_2} & 1 \end{bmatrix}$
  - All loadings estimated
  - $\sigma_{F_1, F_2}$  is now factor correlation

## 2-Factor CFA Model Path Diagram

---





# Disentangling the Model Implied Covariance Matrix

---

For each item measuring a Factor  $f$ :

- Variance of an item  $i$ :  $\lambda_{i,f}^2 \sigma_{F_f}^2 + \Psi_i^2$

For a pair of items measuring **the same** factor  $f$ :

- Covariance of items  $i$  and  $j$ :  $\lambda_{i,f} \lambda_{j,f} \sigma_{F_f}^2$

For a pair of items measuring different factors  $F_1$  and  $F_2$ :

- Covariance of items  $i$  ( $F_1$ ) and  $j$  ( $F_2$ ):  $\lambda_{i,1} \lambda_{j,2} \sigma_{F_1, F_2}$

## Example 2-Factor Model Implied Covariance Matrix

- This covariance matrix is from a 2-factor model
  - Factor 1: I1, I2, I3
  - Factor 2: I4, I5, I6
- Items measuring the same factor are expected to have a high correlation (larger covariances)
- Items measuring different factors are expected (smaller covariances)
  - This is due to the relative size of the covariance between factors – it must be smaller than the factor variance

Test B with 6 items						
Variable	1	2	3	4	5	6
1.	—					
2.	.6	—				
3.	.6	.6	—			
4.	.3	.3	.3	—		
5.	.3	.3	.3	.6	—	
6.	.3	.3	.3	.6	.6	—

# Why Understanding Is Important

---

- Understanding the model implied covariance matrix is the key to:
  - ...making a multifactor model fit well
    - ◆ Using estimated saturated model covariances to help build a model
  - ...forming a hypothesis test to compare relative model fit
    - ◆ Necessary: the model with fewer factors must be nested within the model with more factors to use a likelihood ratio test
  - ...understanding what your model is implying about your data

# Multifactor Model in Mplus

---

```
TITLE:
  Gambling Research Instrument Items
  Data from 1192 College Students/144 Gamblers
  41 Likert Items (1-6): GRI1-GRI41
  12 SOGS items (SOGS4-SOGS15), mostly dichotomous
  =====
  Identification: Marker Item Factor Variance, Zero Factor Mean
  =====
  Two-Factor Model

DATA:
  FILE = alldata_gri.csv;

ANALYSIS:
  ESTIMATOR = MLR;

VARIABLE:
  NAMES = ID GRI1-GRI41;
  USEVARIABLES = GRI1 GRI3 GRI5-GRI6 GRI9-GRI11 GRI13-GRI15 GRI21 GRI23;
  IDVARIABLE = ID;
  MISSING = ALL(99);

MODEL:
  GAMBLING by GRI1 GRI3 GRI5-GRI6 GRI9 GRI10 GRI13 GRI15 GRI21;
  OTHER by GRI11 GRI14 GRI23;

OUTPUT:
  STANDARDIZED MODINDICES(ALL 0) RESIDUAL SAMPSTAT;

SAVEDATA:
  SAVE = FSCORES; !SAVES THE LATENT VARIABLE TRAIT ESTIMATES
  FILE = gri12item_fscores.dat; !PUTS THEM INTO THE FILE NAMED HERE
```

## 2-Factor Model Fit:

- The 2-factor model had good model fit indices:

MODEL FIT INFORMATION		Chi-Square Test of Model Fit	
Number of Free Parameters	37	Value	152.822*
		Degrees of Freedom	53
Loglikelihood		P-Value	0.0000
		Scaling Correction Factor for MLR	1.5961
H0 Value	-18961.922	RMSEA (Root Mean Square Error Of Approximation)	
H0 Scaling Correction Factor for MLR	2.4191	Estimate	0.038
H1 Value	-18839.965	90 Percent C.I.	0.031 0.045
H1 Scaling Correction Factor for MLR	1.9344	Probability RMSEA <= .05	0.998
		CFI/TLI	
		CFI	0.964
		TLI	0.955

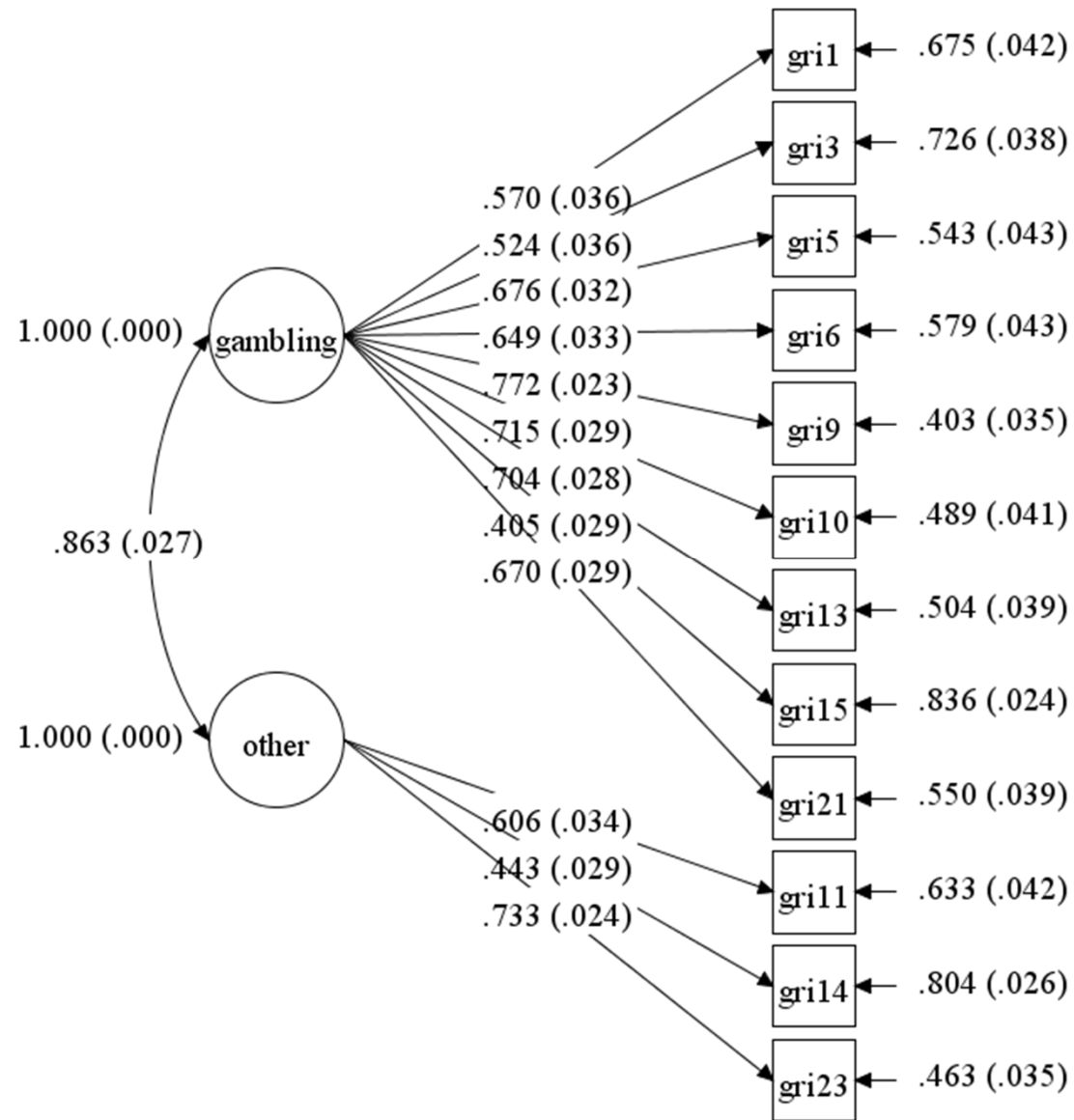
- Note: we cannot tell if this model fits significantly better than the one-factor model (yet)
- The largest normalized residual covariance is now 2.440
  - Between Item 15 and Item 1 – not changed in this model

# 2-Factor Model Interpretation

Unstandardized					Standardized (STDYX)				
MODEL RESULTS					STDYX Standardization				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GAMBLING BY					GAMBLING BY				
GRI1	1.000	0.000	999.000	999.000	GRI1	0.570	0.036	15.639	0.000
GRI3	0.784	0.073	10.725	0.000	GRI3	0.524	0.036	14.405	0.000
GRI5	1.116	0.096	11.595	0.000	GRI5	0.676	0.032	21.147	0.000
GRI6	0.816	0.071	11.533	0.000	GRI6	0.649	0.033	19.541	0.000
GRI9	0.963	0.060	16.179	0.000	GRI9	0.772	0.023	33.835	0.000
GRI10	1.067	0.081	13.156	0.000	GRI10	0.715	0.029	25.052	0.000
GRI13	1.165	0.072	16.138	0.000	GRI13	0.704	0.028	25.128	0.000
GRI15	0.847	0.071	11.976	0.000	GRI15	0.405	0.029	13.789	0.000
GRI21	0.961	0.071	13.568	0.000	GRI21	0.670	0.029	23.331	0.000
OTHER BY					OTHER BY				
GRI11	1.000	0.000	999.000	999.000	GRI11	0.606	0.034	17.556	0.000
GRI14	0.989	0.081	12.251	0.000	GRI14	0.443	0.029	15.329	0.000
GRI23	1.053	0.056	18.925	0.000	GRI23	0.733	0.024	31.002	0.000
OTHER WITH GAMBLING					OTHER WITH GAMBLING				
	0.330	0.036	9.093	0.000		0.863	0.027	31.664	0.000
Variances					R-SQUARE				
GAMBLING	0.335	0.047	7.142	0.000	Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
OTHER	0.437	0.051	8.528	0.000	GRI1	0.325	0.042	7.820	0.000
Residual Variances					GRI3	0.274	0.038	7.202	0.000
GRI1	0.694	0.065	10.603	0.000	GRI5	0.457	0.043	10.574	0.000
GRI3	0.544	0.044	12.351	0.000	GRI6	0.421	0.043	9.770	0.000
GRI5	0.496	0.048	10.427	0.000	GRI9	0.597	0.035	16.918	0.000
GRI6	0.306	0.024	12.810	0.000	GRI10	0.511	0.041	12.526	0.000
GRI9	0.210	0.017	12.597	0.000	GRI11	0.367	0.042	8.778	0.000
GRI10	0.365	0.040	9.190	0.000	GRI13	0.496	0.039	12.564	0.000
GRI11	0.756	0.078	9.689	0.000	GRI14	0.196	0.026	7.664	0.000
GRI13	0.462	0.047	9.866	0.000	GRI15	0.164	0.024	6.894	0.000
GRI14	1.755	0.082	21.365	0.000	GRI21	0.450	0.039	11.665	0.000
GRI15	1.222	0.078	15.650	0.000	GRI23	0.537	0.035	15.501	0.000
GRI21	0.378	0.036	10.525	0.000					
GRI23	0.419	0.039	10.780	0.000					

## 2-Factor Results: Path Diagram of Standardized Coefficients

---



## 2-Factor Interpretation

---

- Because we added our second factor based on statistical misfit, interpretation of what the factors represent is weak at best
  - Must look at the items to figure this out
- Further, the factors were correlated at .863
  - Very high correlation – factors measure virtually the same thing



## Factor 1: Items for Interpretation

---

- The items that measured factor 1:

Item	Criterion	Question
GRI1	3	I would like to cut back on my gambling.
GRI3	6	If I lost a lot of money gambling one day, I would be more likely to want to play again the following day.
GRI5	2	I find it necessary to gamble with larger amounts of money (than when I first gambled) for gambling to be exciting.
GRI6	8	I have gone to great lengths to obtain money for gambling.
GRI9	4	I feel restless when I try to cut down or stop gambling.
GRI10	1	It bothers me when I have no money to gamble.
GRI13	3	I find it difficult to stop gambling.
GRI15	7	I am private about my gambling experiences.
GRI21	1	It is hard to get my mind off gambling.

## Factor 2: Items for Interpretation

---

- The items that measured factor 2:

<u>Item</u>	<u>Criterion</u>	<u>Question</u>
GRI11	5	I gamble to take my mind off my worries.
GRI14	2	I am drawn more by the thrill of gambling than by the money I could win.
GRI23	5	I gamble to improve my mood.

---

# **STATISTICAL METHODS FOR MODEL COMPARISON**

# Comparing Models

---

- We have two well-fitting models for the 12-item GRI
  - One factor
  - Two factors
- We must determine which model fits best (and should be the one reported)
- Model comparisons are made using statistical methods
  - For nested models: likelihood ratio tests (preferred)
  - For non-nested models: comparison of information criteria
- The 1-factor model is nested within the 2-factor model, meaning we can use a likelihood ratio test

# 1-and 2-Factor CFA: Nested Models

---

- To show how the 1-factor model is nested within the 2-factor model, we must look at the model implied variances and covariances
  - We must find a way to make the implied variances/covariances of the 2-factor model match the 1-factor model through setting one or more parameters to a fixed value
- **BIG NOTE**: the following slide is for 2-factor models when items **only** measure one factor (simple structure)
  - *As we will see shortly, items with cross-loadings are more complicated (and need more constraints)*

# Comparing Model Implied Variances and Covariances

Model Portion	1-Factor CFA	2-Factor CFA (w/Simple Structure)
Variance of an item $i$ (measuring factor $f$ )	$\lambda_{i,f}^2 \sigma_{F_f}^2 + \Psi_i^2$	$\lambda_{i,f}^2 \sigma_{F_f}^2 + \Psi_i^2$
Covariance of a pair of items $i$ and $j$ (both measuring the same factor $f$ )	$\lambda_{i,f} \lambda_{j,f} \sigma_{F_f}^2$	$\lambda_{i,f} \lambda_{j,f} \sigma_{F_f}^2$
Covariance of a pair of items $i$ and $j$ (measuring different factors $F_1$ and $F_2$ )	NA	$\lambda_{i,1} \lambda_{j,2} \sigma_{F_1, F_2}$

- Same in both models:
  - Variance of an item  $i$  measuring factor  $f$
  - Covariance of a pair of items  $i$  and  $j$  (both measuring the same factor  $f$ )
- Key to model comparison:
  - Making the covariance of a pair of items  $i$  and  $j$  (measuring different factors  $F_1$  and  $F_2$ ) the same as the variance between two items measuring one factor
  - Can be accomplished if factor **correlation** is set to one (to see, we need standardized factors)

## Standardized Factors: Model Implied Variances and Covariances

Model Portion	1-Factor CFA	2-Factor CFA (w/Simple Structure)
Variance of an item $i$ (measuring factor $f$ )	$\lambda_{i,f}^2 + \Psi_i^2$	$\lambda_{i,f}^2 + \Psi_i^2$
Covariance of a pair of items $i$ and $j$ (both measuring the same factor $f$ )	$\lambda_{i,f}\lambda_{j,f}$	$\lambda_{i,f}\lambda_{j,f}$
Covariance of a pair of items $i$ and $j$ (measuring different factors $F_1$ and $F_2$ )	NA	$\lambda_{i,1}\lambda_{j,2}\sigma_{F_1,F_2}$

- Here  $\sigma_{F_1,F_2}$  is the correlation between the two factors
- If this is set to 1.0, Factor 2 is equal to Factor 1
  - Meaning the loadings for Factor 2 are what they would be onto Factor 1
  - The covariance for items measuring different factors then becomes the product of the factor loadings
  - This is equivalent to a 1-factor model (but only for simple structure tests)

## 2-Factor Model with Standardized Factors

---

- We must first begin by estimating our 2-factor model using standardized factors:

```
MODEL:
  GAMBLING by GRI1* GRI3 GRI5-GRI6 GRI9 GRI10 GRI13 GRI15 GRI21;
  OTHER by GRI11* GRI14 GRI23;

  GAMBLING@1;
  OTHER@1;
```

- We then need the model log-likelihood (for H0), the scaling constant, and the number of parameters

```
Number of Free Parameters          37

Loglikelihood

      H0 Value                      -18961.922
      H0 Scaling Correction Factor    2.4191
      for MLR
      H1 Value                      -18839.965
      H1 Scaling Correction Factor    1.9344
      for MLR
```



## 2-Factor Model with Constraint on Factor Correlation

---

- Next, we need to estimate the 1-factor model
  - In this case, I will use syntax for the 2-factor model with standardized factors and the factor correlation set to 1

```
MODEL:
  GAMBLING by GRI1* GRI3 GRI5-GRI6 GRI9 GRI10 GRI13 GRI15 GRI21;
  OTHER by GRI11* GRI14 GRI23;

  GAMBLING@1;
  OTHER@1;
  GAMBLING WITH OTHER@1;
```

- Again, we need the model log-likelihood and the number of parameters

```
Number of Free Parameters          36

Loglikelihood

      H0 Value                    -18988.425
      H0 Scaling Correction Factor    2.4309
      for MLR
      H1 Value                    -18839.965
      H1 Scaling Correction Factor    1.9344
      for MLR
```

# Model Comparison Likelihood Ratio Test

---

- We now have all the information we need to conduct a likelihood ratio test comparing the 2-factor model to the 1-factor model
  - $H_0$ : 1-factor model fits data (nested model);  $\sigma_{F_1, F_2} = 1$
- The likelihood ratio test statistic is  $-2 * (\text{difference in log-likelihoods})$ :
$$\chi^2 = -2 * (-16,648.054 - -16,621.205) / 1.9943 = 26.579$$
  - The test has one degree of freedom (the difference in the number of parameters)
- If we use a standard Chi-Square distribution, the p-value is  $< .001$  – therefore we reject the 1-factor model in favor of the 2-factor model
  - But this test is not the right one here...

## Additional Complication: Model Comparison Occurs on a Boundary

---

- The likelihood ratio test from the previous slide was attained by setting the correlation between factors to 1
  - This is the boundary of the correlation
- Parameters fixed at a boundary violate one of the conditions for use of the likelihood ratio test statistic
  - The violation means the distribution of the test statistic is not a standard  $\chi^2$  with 1-df
- Instead, the null distribution is a mixture of  $\chi^2$  distributions:  $.5\chi_0^2 + .5\chi_1^2$
- In practice this is ignored – rejections of the null hypothesis will be extremely conservative
  - Meaning parsimony is heavily favored

# But About That 1-Factor Model

---

- The log-likelihood under the constrained 2-factor model:

```
Number of Free Parameters          36

Loglikelihood

      H0 Value          -18988.425
      H0 Scaling Correction Factor    2.4309
      for MLR
      H1 Value          -18839.965
      H1 Scaling Correction Factor    1.9344
      for MLR
```

- The log-likelihood under the original 1-factor model:

```
Number of Free Parameters          36

Loglikelihood

      H0 Value          -18988.425
      H0 Scaling Correction Factor    2.4309
      for MLR
      H1 Value          -18839.965
      H1 Scaling Correction Factor    1.9344
      for MLR
```

- As they are identical we could have just constructed the LR test with the original parameterization
  - However, this obscures the process a little...especially with items that have cross-loadings

---

## **CROSS LOADINGS: ITEMS MEASURING MORE THAN ONE FACTOR**

# Re-examining the 2-Factor Model

---

- Thus far we have concluded that the 2-factor (simple structure) model fit better than 1-factor model
- Examination of the modification indices from the 2-factor model indicated a large modification index for one additional factor loading:

	M.I.	E.P.C.	Std E.P.C.	StdYX E.P.C.
ON/BY Statements				
GRI23 ON GAMBLING /				
GAMBLING BY GRI23	9.520	0.789	0.789	0.830

- This MI is for adding the loading of GRI23 onto the first factor – making it load onto both
  - This is called a cross-loaded item
- I will add this cross-loading so as to show how to build a LRT between 1- and 2-factor models

# Cross-Loaded Items: Conceptual Complications

---

- Cross-loaded items have traditionally been thought of in disparaging terms in the CFA literature
  - It has been thought that items are “bad” if they measure more than one factor
- Conceptually, it is a matter of the grain-size of the latent trait – very coarse level traits are hard to envision being measured by multiple items
- This thinking is slowly starting to change as knowledge of latent traits becomes deeper
  - In educational tests – certain constructs must load onto more than one item: Diagnostic classification models are built under this assumption
  - Method factors in CFA often use these types of items
  - Bi-factor models in CFA use these types of items, too

# Cross-Loaded Items: Statistical Complications

---

- In addition to whether or not an item can load onto more than one trait conceptually, cross-loaded items present a few more statistical challenges
  - It is harder to compare a 1-factor model to a multi-factor model with cross-loaded items
  - Some of the loadings must be set to zero in addition to the factor correlation to one
- To demonstrate, we will add the loading of GRI23 onto Factor 1 and compare the model fit
  - The comparison will be against the 1-factor model
  - Assume the process of analysis lead to the 2-factor model with the cross-loading as the candidate for the final model



## 2-Factor Model with Cross Loadings

- Using the standardized factors identification:

```
MODEL:
    GAMBLING by GRI1* GRI3 GRI5-GRI6 GRI9 GRI10 GRI13 GRI15 GRI21 GRI23;
    OTHER by GRI11* GRI14 GRI23;

    GAMBLING@1;
    OTHER@1;
```

Number of Free Parameters 38

Loglikelihood

H0 Value	-18954.481
H0 Scaling Correction Factor for MLR	2.4056
H1 Value	-18839.965
H1 Scaling Correction Factor for MLR	1.9344

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.037
90 Percent C.I.	0.030 0.044
Probability RMSEA <= .05	0.999

CFI/TLI

CFI	0.966
TLI	0.957

Chi-Square Test of Model Fit

Value	144.033*
Degrees of Freedom	52
P-Value	0.0000
Scaling Correction Factor for MLR	1.5901

# Statistical Issue for Model Comparison: Additional Constraints

---

- Because of the cross-loaded item, we cannot simply set the factor correlation to 1 to compare the fit of the model to a 1-factor model
  - More terms must be set to zero
- Under the standardized factor identification, for items with multiple loadings (here onto 2 factors)

- Implied variance of cross-loaded item  $i$ :

$$\lambda_{i,F_1}^2 + \lambda_{i,F_2}^2 + 2\lambda_{i,F_1}\lambda_{i,F_2} + \psi_i^2$$

- Implied covariance of cross-loaded item  $i$  with non-cross-loaded item  $j$ :

$$\lambda_{i,F_1}\lambda_{j,F_1} + \lambda_{i,F_2}\lambda_{j,F_1}$$

- What this means: we must set the cross-loadings to 0 in addition to the factor correlation to 1 to have the 1-factor model as a nested model
  - Practically speaking, you can still just compare log-likelihoods from the original model output

# The Nested Model in Mplus

---

- Phrasing the 2-factor model as a 1-factor:

```
MODEL:
  GAMBLING by GRI1* GRI3 GRI5-GRI6 GRI9 GRI10 GRI13 GRI15 GRI21 GRI23@0;
  OTHER by GRI11* GRI14 GRI23;

  GAMBLING@1;
  OTHER@1;

  GAMBLING WITH OTHER@1;
```

- Log-likelihood (same as from 1-factor model):

Number of Free Parameters	36
Loglikelihood	
H0 Value	-18988.425
H0 Scaling Correction Factor for MLR	2.4309
H1 Value	-18839.965
H1 Scaling Correction Factor for MLR	1.9344

- Null hypothesis:  $\sigma_{F_1, F_2} = 1$  **and**  $\lambda_{23,2} = 0$  (simultaneously)
- MLR Likelihood ratio test:  $\chi^2 = 34.811$  (2 DF now),  $p < .001$ 
  - Again this is a non-standard test (very conservative)

# Model Parameter Interpretation

---

- The parameters of the model are interpreted similarly:

## STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GAMBLING BY				
GRI1	0.570	0.036	15.638	0.000
GRI3	0.524	0.036	14.386	0.000
GRI5	0.676	0.032	21.090	0.000
GRI6	0.649	0.033	19.490	0.000
GRI9	0.772	0.023	33.758	0.000
GRI10	0.714	0.029	25.025	0.000
GRI13	0.705	0.028	25.180	0.000
GRI15	0.406	0.029	13.809	0.000
GRI21	0.671	0.029	23.314	0.000
GRI23	0.389	0.088	4.434	0.000
OTHER BY				
GRI11	0.686	0.038	18.234	0.000
GRI14	0.488	0.033	14.835	0.000
GRI23	0.338	0.092	3.669	0.000

---

# METHOD FACTORS

# Method Factors

---

- Method factors are latent constructs that summarize variability due to differing testing methods
  - Most common in use in negatively worded items
- Method factors are additional latent variables that are measured by items with features that are similar
  - Such as our negatively worded items
- The method factors are typically uncorrelated with the factor of interest
  - The “method” items are then cross loaded onto each

# Method Factors and Our Data

---

- Recall that the GRI had three reverse scored items
  - Each of these fit the 1-factor model very poorly
- We can attempt to add these to our analysis by using a method factor for the three items
  - Method = Negatively worded items
  - Idea – that the response to these items is influenced by not only the item but the method in which it is displayed
- The method factor will be uncorrelated with our general gambling factor

# Mplus Syntax and Model Fit Information

- The Mplus syntax:

```
MODEL:
    GAMBLING by GRI1 GRI3 GRI5-GRI6 GRI9 GRI10 GRI13 GRI15 GRI21 GRI11 GRI14 GRI23
                GRI4 GRI12 GRI20;
    NEGATIVE by GRI4 GRI12 GRI20;

    GAMBLING WITH NEGATIVE@0;
```

- The model fit information:

## MODEL FIT INFORMATION

Number of Free Parameters	48
Loglikelihood	
H0 Value	-26225.191
H0 Scaling Correction Factor for MLR	2.0737
H1 Value	-25861.459
H1 Scaling Correction Factor for MLR	1.6618

## Chi-Square Test of Model Fit

Value	507.127*
Degrees of Freedom	87
P-Value	0.0000
Scaling Correction Factor for MLR	1.4345

## RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.061
90 Percent C.I.	0.056 0.066
Probability RMSEA <= .05	0.000

## CFI/TLI

CFI	0.894
TLI	0.872

- We will claim this is adequate model fit – and continue with model comparison (to the 1-factor model)



# Model Comparison

---

- Because the method factor model fit “adequately,” we will now compare the relative fit of the method factor model to the 1-factor model
- The method factor model had 15 items – so we cannot use the 12 item 1-factor model as our comparison
  - Models with differing sets of items are NOT nested so cannot be compared with a LRT
- We must run a 15 item 1-factor model:

MODEL:

```
GAMBLING by GRI1 GRI3 GRI5-GRI6 GRI9 GRI10 GRI13 GRI15 GRI21 GRI11 GRI14 GRI23  
GRI4 GRI12 GRI20;
```

# Model Comparison

---

- The log-likelihoods from our models:

## MODEL FIT INFORMATION

Number of Free Parameters 45

## Loglikelihood

H0 Value	-26507.144
H0 Scaling Correction Factor for MLR	2.1092
H1 Value	-25861.459
H1 Scaling Correction Factor for MLR	1.6618

## MODEL FIT INFORMATION

Number of Free Parameters 48

## Loglikelihood

H0 Value	-26225.191
H0 Scaling Correction Factor for MLR	2.0737
H1 Value	-25861.459
H1 Scaling Correction Factor for MLR	1.6618

- The null hypothesis: that the method loadings are all simultaneously zero (not a boundary)
- MLR Test statistic:  $\chi^2 = 365.888$ , DF = 3;  $p < .001$
- Therefore, the method factor model fits better than the 1-factor model for these data

# Other Models with Uncorrelated Factors

---

- The method factor model added an uncorrelated second factor to the analysis
- Another model with an uncorrelated factor is the bifactor model – one general factor measured by all items and sub-factors measured by subsets of items
  - The general factor is uncorrelated with all sub factors
  - The subfactors are all intercorrelated
- The bifactor model originated in examining intelligence data – one overall general ability and specific content area sub abilities

---

# **BUILDING A MULTI-FACTOR SCALE**

# Building a Multi-Factor Scale

---

- Our previous 2-factor model was built from statistical information rather than substantive theory
  - Not a great idea in practice
- To demonstrate multifactor CFA, we will now use all 41 items of the GRI to build a 10-factor gambling model
  - Each factor represents a criterion from the DSM
- Along the way, this will allow us to introduce a few additional concepts in multi-factor CFA analyses

# Important Concepts in Multifactor CFA

---

- The multifactor CFA model presents new complexities we must be able to handle appropriately:
  - Model estimation can go badly
    - ◆ Many more types of estimation errors
  - Model modification may be difficult
    - ◆ Simultaneous analysis may be impossible
    - ◆ Item removal may lead to removal of factors
  - Model fit now must consider covariance matrix of factors
    - ◆ Reduced versions of these exist (CFA for the factors) called structural models – but each will make fit worse compared to saturated model for factor covariances

## Full 41 Item GRI: Constructed to Measure 10 Factors

---

- The 41 items of the GRI were constructed to measure a 10-factor model (one factor for each DSM criterion)
  - Note: one cross-loaded item (item 39 is thought to measure criterion 3 and criterion 6)

<u>Criterion</u>	<u>Item Count</u>
1	6
2	3
3	8
4	2
5	4
6	5
7	3
8	4
9	3
10	4

# 10 Factor Model Analysis Process

---

- We will begin by attempting to estimate the entire 10-factor model first
  - We will use an saturated factor covariance matrix...and quickly find that this is impossible to estimate
- We will then look at each factor **separately** and attempt to make a 1-factor model fit for each factor
  - If items do not fit well within a single factor, they will not fit well with other items across the analysis



# Full Model Analysis in Mplus

---

- The 10-Factor 41 item model:

```
MODEL:
  C1 BY GRI2 GRI10 GRI21 GRI26 GRI34 GRI41;
  C2 BY GRI5 GRI14 GRI27;
  C3 BY GRI1 GRI8 GRI13 GRI19 GRI24 GRI29 GRI36 GRI39;
  C4 BY GRI9 GRI31;
  C5 BY GRI7 GRI11 GRI17 GRI23;
  C6 BY GRI3 GRI32 GRI39 GRI12 GRI20;
  C7 BY GRI15 GRI35 GRI4;
  C8 BY GRI6 GRI16 GRI28 GRI37;
  C9 BY GRI18 GRI30 GRI38;
  C10 BY GRI22 GRI25 GRI33 GRI40;
```

- The Mplus output starts with a warning:

```
THE MODEL ESTIMATION TERMINATED NORMALLY
```

```
WARNING:  THE LATENT VARIABLE COVARIANCE MATRIX (PSI) IS NOT POSITIVE
DEFINITE.  THIS COULD INDICATE A NEGATIVE VARIANCE/RESIDUAL VARIANCE FOR A
LATENT VARIABLE, A CORRELATION GREATER OR EQUAL TO ONE BETWEEN TWO LATENT
VARIABLES, OR A LINEAR DEPENDENCY AMONG MORE THAN TWO LATENT VARIABLES.
CHECK THE TECH4 OUTPUT FOR MORE INFORMATION.
PROBLEM INVOLVING VARIABLE C4.
```

# Non-Positive Definite Factor Covariance Matrices

---

- The warning from Mplus regards the covariance matrix for the factors (not the data, the factors)
- The term non-positive definite means this matrix is not invertible – which further means that the model implied covariance matrix for the items will not be non-positive definite as well
  - NPD = not invertible (so a more general inverse is used)
- When you see an error like this, it means your model isn't working the way you thought it should
  - So don't interpret the results
- Using the STDYX standardization, here are some correlations amongst factors:
  - Correlations bigger than 1 aren't possible in reality

C7	WITH				
C1		1.260	0.086	14.706	0.000
C2		1.163	0.094	12.437	0.000
C3		0.879	0.214	4.112	0.000
C4		0.991	0.201	4.939	0.000
C5		1.114	0.114	9.730	0.000
C6		1.362	0.096	14.256	0.000

# Where to Go Next

---

- Perhaps a better place to go would be to examine each factor separately – making sure a 1-factor model fit for each factor
- Recall the minimum number of items for a factor:
  - 2: with other factors in a model – not for separate analyses
  - 3: just identified - model fit is perfect (so cannot assess)
  - 4: over identified – can assess model fit
- Our 10-factor analysis had 4 factors with measured by 2 or 3 items (C2, C4, C7, and C9) – we cannot assess these
  - These are likely causing the problems in estimation – but we cannot do much about that now (should have had more items)

# One Factor Models

Criterion	Initial Number of Items	Initial RMSEA/CFI/TLI	Final Number of Items	Final RMSEA/CFI/TLI
C1	6	.108/.890/.817	4	.007/1.000/1.000
C2	3		3	
C3	8	.091/.839/.774	5	.058/.966/.933
C4	2		2	
C5	4	.064/.978/.935	4	.064/.978/.935
C6	5	.184/.667/.333	4	.015/.998/.995
C7	3		3	
C8	4	.034/.990/.970	4	.034/.990/.970
C9	3		3	
C10	4	.210/.739/.217	3	

# New 10-Factor Analysis

---

- Using our reduced item sets, we now reattempt to estimate the 10-factor GRI:

MODEL:

```
C1 BY GRI2 GRI10 GRI26 GRI41;  
C2 BY GRI5 GRI14 GRI27;  
C3 BY GRI1 GRI8 GRI13 GRI19 GRI39;  
C4 BY GRI9 GRI31;  
C5 BY GRI7 GRI11 GRI17 GRI23;  
C6 BY GRI3 GRI32 GRI39 GRI20;  
C7 BY GRI15 GRI35 GRI4;  
C8 BY GRI6 GRI16 GRI28 GRI37;  
C9 BY GRI18 GRI30 GRI38;  
C10 BY GRI22 GRI25 GRI33;
```

- But we get the same error:

```
THE MODEL ESTIMATION TERMINATED NORMALLY
```

```
WARNING:  THE LATENT VARIABLE COVARIANCE MATRIX (PSI) IS NOT POSITIVE  
DEFINITE.  THIS COULD INDICATE A NEGATIVE VARIANCE/RESIDUAL VARIANCE FOR A  
LATENT VARIABLE, A CORRELATION GREATER OR EQUAL TO ONE BETWEEN TWO LATENT  
VARIABLES, OR A LINEAR DEPENDENCY AMONG MORE THAN TWO LATENT VARIABLES.  
CHECK THE TECH4 OUTPUT FOR MORE INFORMATION.  
PROBLEM INVOLVING VARIABLE C4.
```

# Options for Analysis

---

- The data is trying to tell us we do not have 10 factors!
- At this point we have a few options for how to proceed:
  - Add stuff up for each Criterion, then use in a factor analysis
    - ♦ Cons:
      - Dangerous! Parceling (but with fit check for some criteria)
      - Will hid model misfit
      - Therefore, I am not encouraging this behavior...
    - ♦ Pros:
      - Will allow for all 10 “factors” to be used (but within each factor we cannot assess fit)
      - Similar to factor structural model
  - Remove some factors (and their items) from the analysis
    - ♦ Cons:
      - Changes the nature of the construct
      - May not result in model that works
      - Can be very frustrating
    - ♦ Pros:
      - Mplus indicates the problem is with C4 – so we can start there
      - Can assess model fit (if we can get it to work)

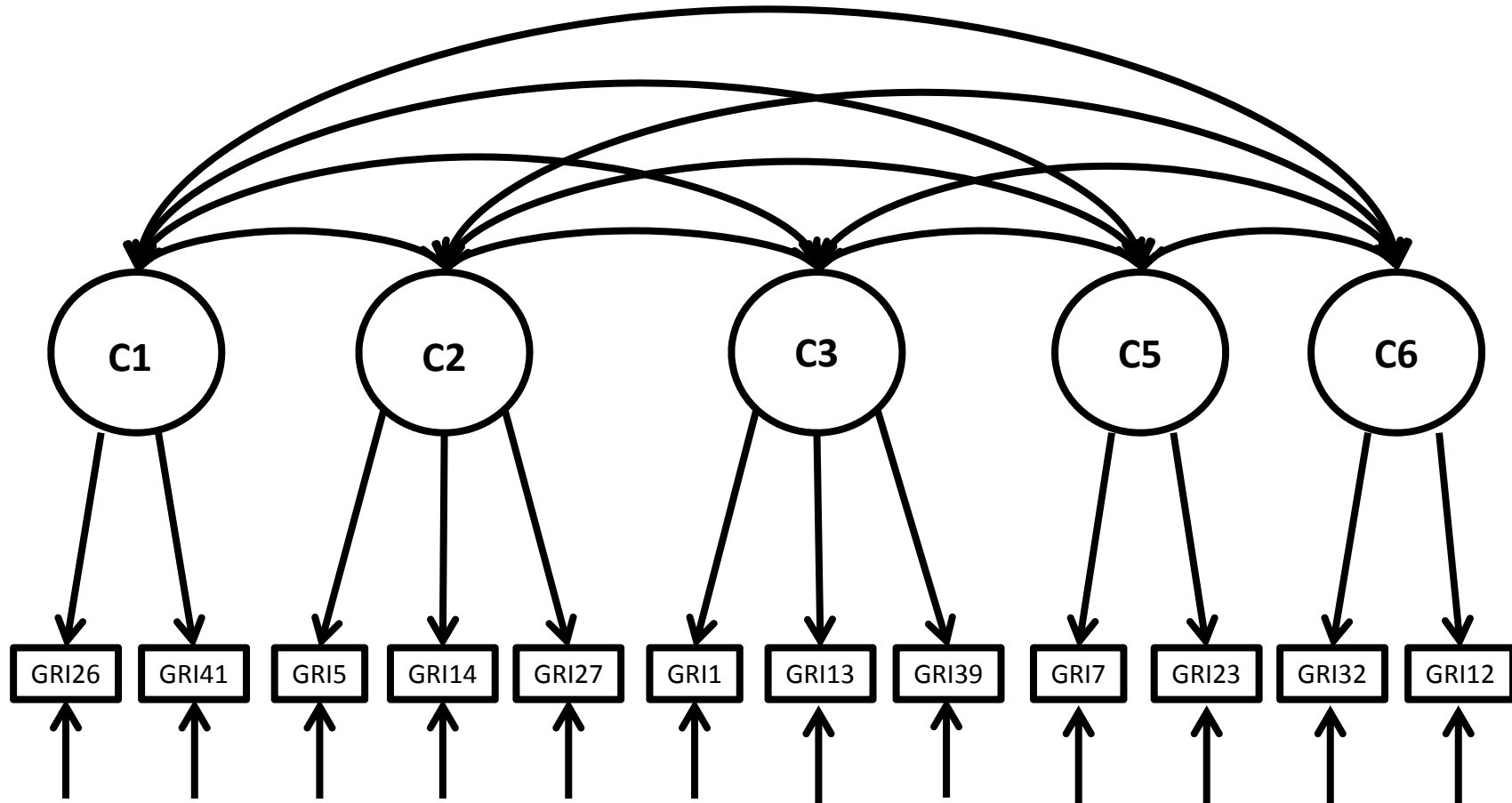
# Removing Factors from the Analysis

---

- Mplus told us it suspected an issue with factor C4
  - The factor was measured by only 2 items (the minimum)
- It was removed, as was factors C7, C8, and C9
  - Then the model worked...but fit very poorly
- So then.....
  - .....omitted so you won't follow these steps
  - If you are in this case, question your test items and question your theory!
- And in the end, I came up with a 5-factor model that fit adequately enough to give you an example

# The 5-Factor Model

---





# Mplus Syntax and Model Fit Information

## TITLE:

```
Gambling Research Instrument Items
Data from 1192 College Students/144 Gamblers
41 Likert Items (1-6): GRI1-GRI41
12 SOGS items (SOGS4-SOGS15), mostly dichotomous
=====
Identification: Marker Item Factor Variance, Zero Factor Mean
=====
All 41 Items: 10 Criteria Factors
```

## DATA:

```
FILE = gamblingdata.csv;
```

## ANALYSIS:

```
ESTIMATOR = MLR;
ITERATIONS = 100000;
```

## VARIABLE:

```
NAMES = GRI1-GRI41 SOGS4-SOGS15 Student ID;
USEVARIABLES = GRI26 GRI41
                GRI5 GRI14 GRI27
                GRI1 GRI8 GRI13
                GRI7 GRI23
                GRI12 GRI32;
IDVARIABLE = ID;
MISSING = ALL(99);
```

## MODEL:

```
C1 BY GRI26 GRI41;
C2 BY GRI5 GRI14 GRI27;
C3 BY GRI1 GRI8 GRI13 GRI39;
C5 BY GRI7 GRI23;
C6 BY GRI12 GRI32 GRI39;
```

## OUTPUT:

```
STANDARDIZED MODINDICES(ALL 0) RESIDUAL SAMPSTAT;
```

## MODEL FIT INFORMATION

Number of Free Parameters	49
---------------------------	----

### Chi-Square Test of Model Fit

Value	271.406*
Degrees of Freedom	55
P-Value	0.0000
Scaling Correction Factor	1.3416
for MLR	

### RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.054
90 Percent C.I.	0.048 0.061
Probability RMSEA <= .05	0.132

### CFI/TLI

CFI	0.949
TLI	0.928

### SRMR (Standardized Root Mean Square Residual)

Value	0.032
-------	-------

## Before Interpretation: Examine if 5 Factor Model Fits Better Than 1 Factor Model

---

- Before using the 5 Factor Model, we must see if it fits better than a 1 factor model using the same items
- As these items are different from the 1 factor model we built earlier in class, we must estimate the 1 factor model on these items
  - Models with different items cannot be nested

From Mplus (1 factor model):

```
MODEL FIT INFORMATION

Number of Free Parameters      39
Number of Free Parameters      39

Loglikelihood

      H0 Value              -23325.158
      H0 Scaling Correction Factor    1.8926
      for MLR
```

From Mplus (5 factor model):

```
MODEL FIT INFORMATION

Number of Free Parameters      49

Loglikelihood

      H0 Value              -23210.018
      H0 Scaling Correction Factor    1.9468
      for MLR
```

- The likelihood ratio test essentially tests the following constraints, simultaneously:
  - All 10 factor correlations = 1
- LRT Results:  $\chi^2_{MLR} = 106.701$ , DF = 10,  $p < .001$
- Therefore, we will use the 5 factor model

# 5-Factor Model Standardized Results: Measurement Model Parameters

STDYX Standardization

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value	R-SQUARE				
						Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
C1	BY									
	GRI26	0.897	0.015	61.308	0.000	GRI26	0.805	0.026	30.654	0.000
	GRI41	0.800	0.018	45.391	0.000	GRI41	0.639	0.028	22.695	0.000
C2	BY					GRI5	0.493	0.035	14.167	0.000
	GRI5	0.702	0.025	28.334	0.000	GRI14	0.170	0.022	7.713	0.000
	GRI14	0.412	0.027	15.427	0.000	GRI27	0.684	0.031	22.033	0.000
	GRI27	0.827	0.019	44.065	0.000	GRI1	0.387	0.048	8.061	0.000
C3	BY					GRI8	0.353	0.045	7.777	0.000
	GRI1	0.622	0.039	16.122	0.000	GRI13	0.467	0.040	11.620	0.000
	GRI8	0.594	0.038	15.555	0.000	GRI39	0.442	0.044	10.060	0.000
	GRI13	0.684	0.029	23.239	0.000	GRI7	0.384	0.038	10.194	0.000
						GRI23	0.500	0.037	13.562	0.000
C5	BY					GRI12	0.025	0.008	3.150	0.002
	GRI7	0.620	0.030	20.388	0.000	GRI32	0.532	0.034	15.442	0.000
	GRI23	0.707	0.026	27.125	0.000					
C6	BY									
	GRI12	0.159	0.025	6.300	0.000					
	GRI32	0.729	0.024	30.884	0.000					
	GRI39	0.665	0.033	20.121	0.000					

## 5-Factor Model Standardized Results: Structural Model Parameters

---

- The factor correlations (found under STDYX output):

C2	WITH				
C1		0.875	0.021	42.127	0.000
C3	WITH				
C1		0.751	0.030	25.409	0.000
C2		0.819	0.039	20.834	0.000
C5	WITH				
C1		0.940	0.032	29.210	0.000
C2		0.925	0.040	23.215	0.000
C3		0.904	0.044	20.707	0.000
C6	WITH				
C1		0.873	0.027	31.931	0.000
C2		0.911	0.032	28.597	0.000
C3		0.863	0.031	27.660	0.000
C5		0.909	0.038	23.839	0.000

- Arranged more compactly (more on this shortly):

$$\Phi = \begin{bmatrix} 1 & & & & \\ .875 & 1 & & & \\ .751 & .819 & 1 & & \\ .940 & .925 & .904 & 1 & \\ .873 & .911 & .863 & .909 & 1 \end{bmatrix}$$

## Interpreting the Results: Each Factor

---

- Factor C1: Preoccupation with Gambling
  - Item 26: **I think about gambling.**
  - Item 41: **I think about my past gambling experiences.**
- Factor C2: Gambles with Increasing Amounts for Excitement
  - Item 5: **I find it necessary to gamble with larger amounts of money (than when I first gambled) for gambling to be exciting.**
  - Item 14: **I am drawn more by the thrill of gambling than by the money I could win.**
  - Item 27: **I make larger bets than I did when I first started gambling.**

## Interpreting the Results: Each Factor

---

- Factor C3: Unsuccessful Efforts to Control Gambling
  - Item 1: **I would like to cut back on my gambling.**
  - Item 13: **I find it difficult to stop gambling.**
- Factor C5: Gambles for Escape
  - Item 7: **I feel “high” when I gamble.**
  - Item 23: **I gamble to improve my mood.**
- Factor C6: Chases Losses
  - Item 12: **When I lose money gambling, it is a long time before I gamble again. (Reverse Coded)**
  - Item 32: **After losing money, I gamble again to win it back.**
  - Item 39: **I have spent more money gambling than I intended to.**

# Common Methods for Scoring Multifactor Tests

---

- Now that we found good fit for a 5 factor model, a next step is that of using the test
  - And giving subjects test scores
- With multidimensional factor models, a score for each factor should be given
  - Along with a reported reliability and standard error
  - Can be done with a sum score (but use Omega reliability)
- What is often also reported is one single score for the entire test
  - This defeats the purpose of the 5 factor analysis (so I won't report it)

# Omega Reliability Coefficients for Sum Scores

---

- Factor C1: Preoccupation with Gambling
  - Omega = .843
- Factor C2: Gambles with Increasing Amounts for Excitement
  - Omega = .648
- Factor C3: Unsuccessful Efforts to Control Gambling
  - Omega = .661
- Factor C5: Gambles for Escape
  - Omega = .606
- Factor C6: Chases Losses
  - Omega = .380



# What We Have Learned So Far

---

- The 10-Factor Model did not work with this test
  - Had to use fewer factors – had 5
- The 5-factor model fit better than a 1-factor model
  - Most items had high  $R^2$  - except for items 12 and 14
- The reliability of the scale subscores is low
  - No items measuring each – not a good test to administer
- Any analysis based on this test would have low power
  - But results may be a good place to start to build a better test

# One More Thing

---

- In the process of removing poorly fitting items, when a factor had less than two items, the factor was removed (meaning the other item was dropped)
- There are some that would include these single items
  - Sometimes as single indicator factors (but what is the reliability of a single item?)
  - Sometimes left to correlate directly with other factors
- This is not an effective practice as the item will likely contribute more to model misfit than to model fit
  - And any inclusion will not add to results

---

# **FACTOR STRUCTURAL MODELS:**

# Factor Structural Models

---

- The saturated factor covariance matrix  $\Phi$ , although used in building a multi-factor CFA model (the measurement model), is often not of interest
- More commonly, theories about the nature of the factors exists –the factor covariances are where theories get tested
- We will now try a common factor structural model to demonstrate the process
- Of note:
  - The saturated factor covariance matrix will fit the best
  - Estimation is simultaneous – we add this once we have a well-fitting measurement model
  - The model is for the factors themselves

# Understanding Factor Structural Models

---

- The key to understanding structural model for the factors is to treat the factors as observed variables
  - The structural model uses relationships between the factors
  - The overall model parcels measurement error from factor variance giving a more powerful and less biased estimate
- Some models can be only based on the factors
  - Higher-order factor model (all factors then load onto a higher-level overarching set of factors)
- Some can be with external observed variables
  - These are path models with latent variables
- Models can be combinations of both
  - Hierarchical factors that predict observed variables (or other factors)

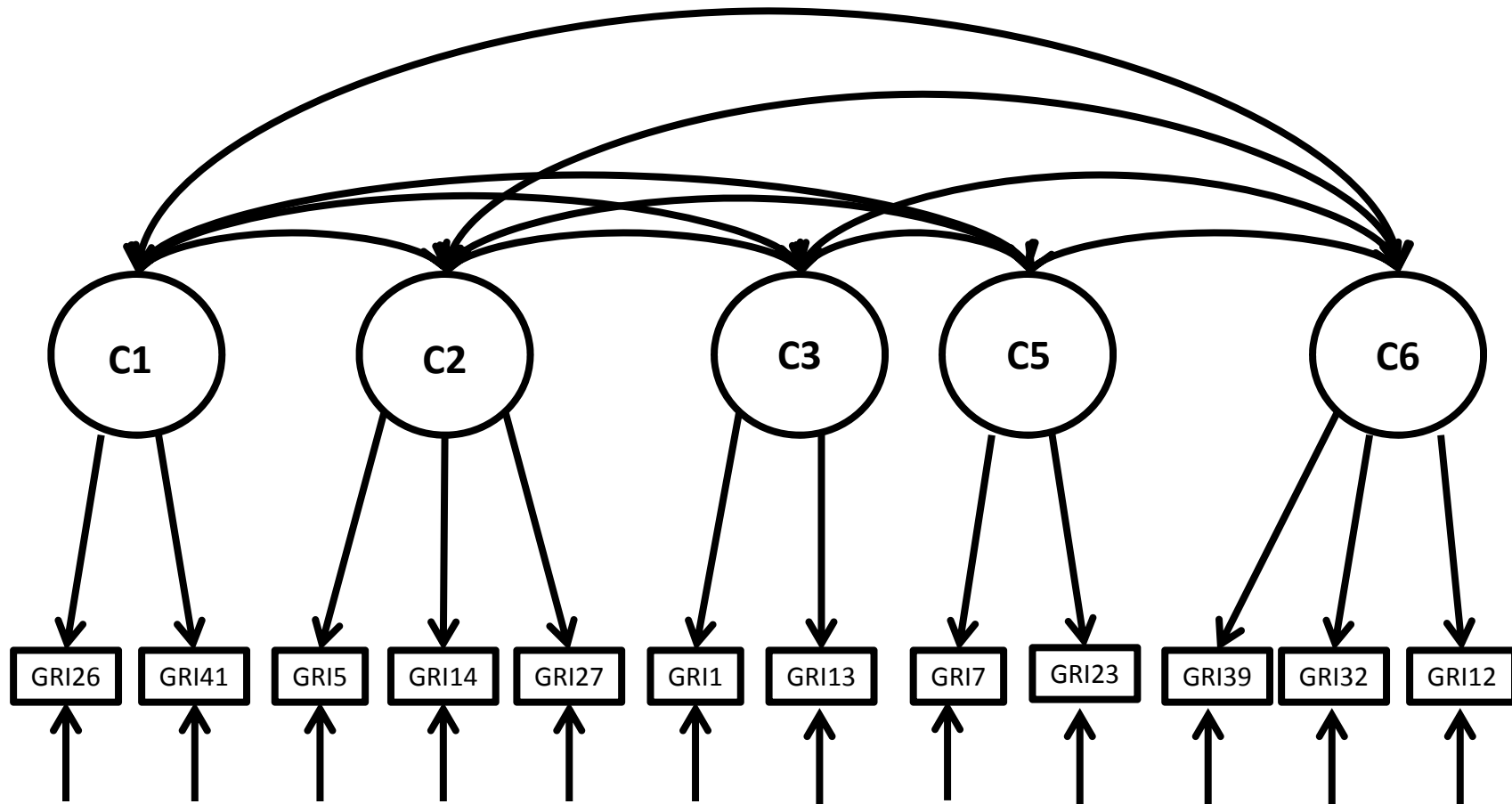
# Higher-Order Factor Model

---

- The higher order factor model states that a general higher-level factor is measured by the first order factors
  - Factor analysis of the factors
- As with the set of equations in a measurement model, the structural model has a set of equations for the factors
- To demonstrate, we will now test the following two higher order factor models
  - A 1-factor higher order model:
    - ◆ All factors measure a general gambling higher-order factor
  - A 2-factor higher order model:
    - ◆ Factors C1 and C2 measure a higher order dependency factor where as factors C3, C5, and C6 measure a higher order loss of control factor

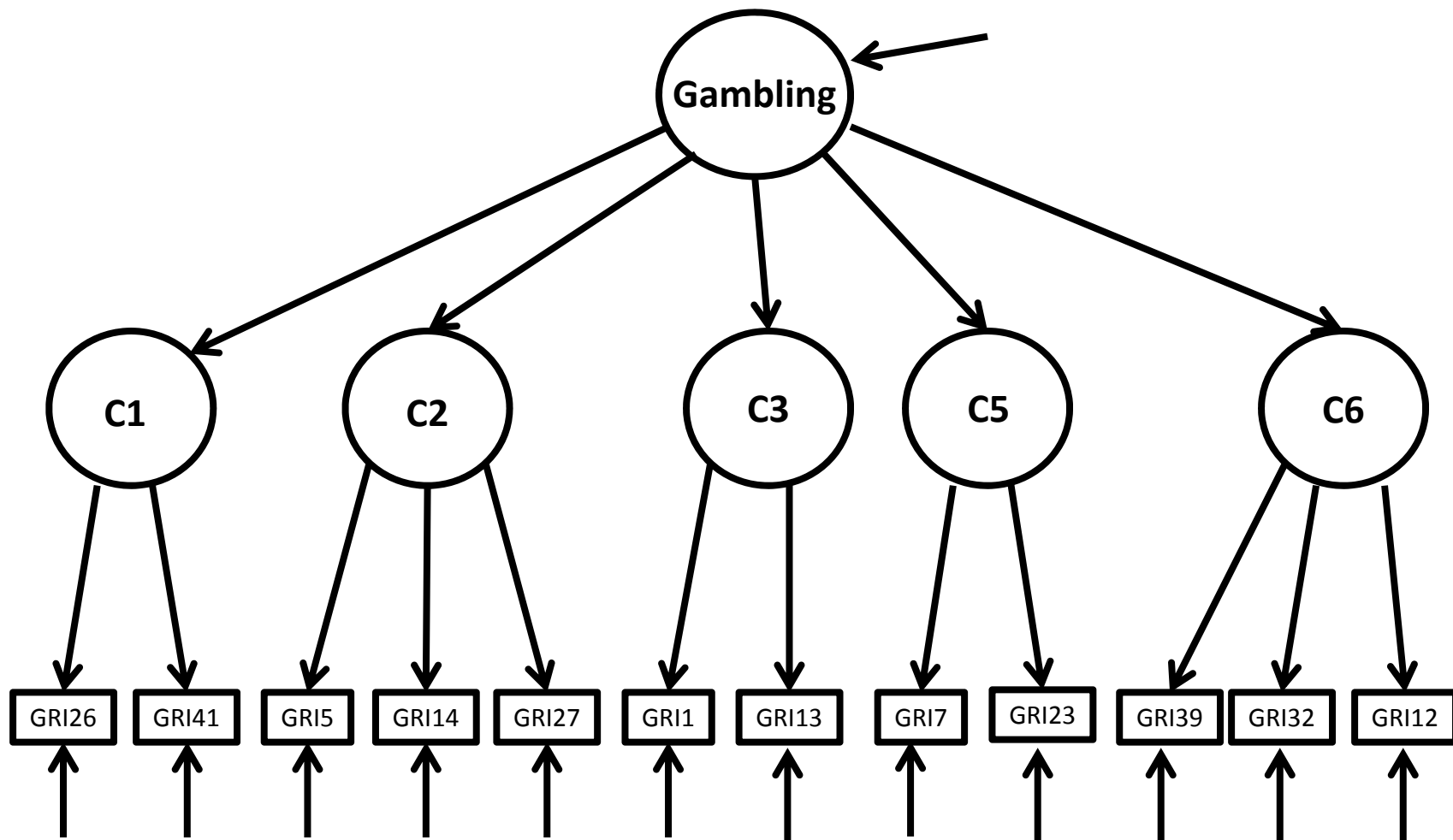
## The 5-Factor Model: Saturated Factor Covariances

---



## The 5-Factor Model: One Higher Order Factor

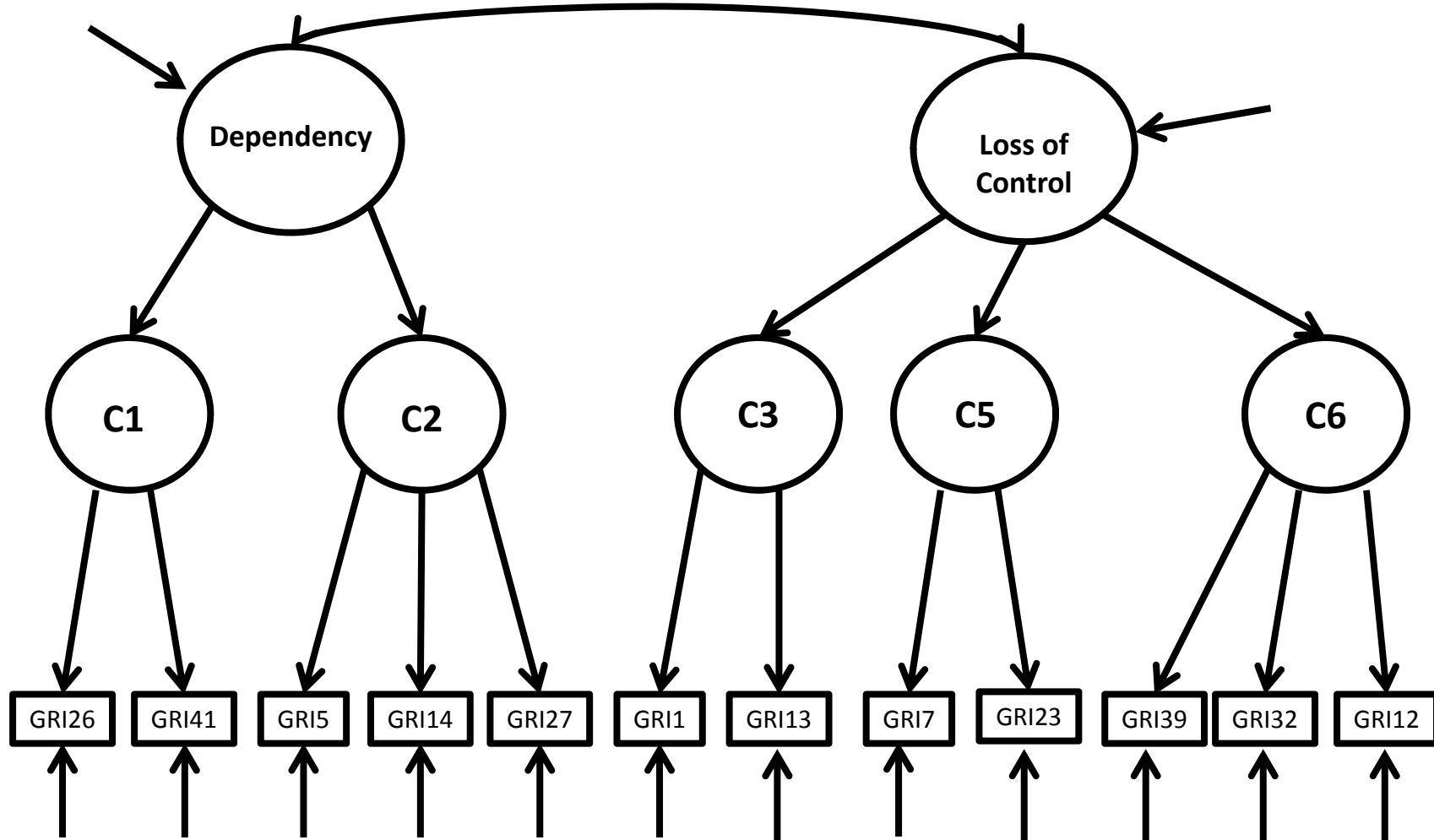
---





# The 5-Factor Model: Two Higher Order Factors

---



# Key Modeling Issues

---

- As with the measurement model, the factor structural model is constrained by the saturated version
  - Cannot have more model parameters than factor variances and covariances
  - With 5-factors we have 15 unique terms in our factor covariance matrix
- Because structural models often change factor variances, the marker item method generally works best
- Model comparisons function the same way
  - For nested models: likelihood ratio tests
  - For non-nested models: comparisons of information criteria

# Higher Order Factor Model In Equations

---

- Factor equations take the form of regression equations:

$$F_{s,f} = \gamma_{I_{C1}} + \xi_{C1,1}G_s + \cdots + u_{sC1}$$

- Where  $G$  is the higher order factor score(s) (assumed  $N(0, \mathbf{P})$ )
- $u$  is the higher order residual (assumed  $N(0, \mathbf{U})$ )

- This has implications for the factor covariance matrix:

$$\mathbf{\Phi} = \mathbf{\Xi P \Xi}^T + \mathbf{U}$$

- The model implied covariance matrix for the observed data is then:

$$\mathbf{\Sigma} = \mathbf{\Lambda \Phi \Lambda}^T + \mathbf{\Psi} = \mathbf{\Lambda(\Xi P \Xi}^T + \mathbf{U)\Lambda}^T + \mathbf{\Psi}$$

# Baseline Model: Saturated Factor Covariances

---

- Our measurement model provides the best fitting model – and is the model to which each higher order model is compared
- We will first construct the 2 higher order factors model
  - If it does not fit better than saturated model we can stop
  - 1 higher order factor model is more restrictive

```
MODEL:
  C1 BY GRI26 GRI41;
  C2 BY GRI5 GRI14 GRI27;
  C3 BY GRI1 GRI8 GRI13;
  C5 BY GRI7 GRI23;
  C6 BY GRI12 GRI32 GRI39;

  DEPEND BY C1 C2;
  LOC BY C3 C5 C6;
```

## 2 Higher Order Factors: Mplus Syntax and Output

---

- The Mplus syntax treats the factors as if they were observable variables:

```
MODEL:
  C1 BY GRI26 GRI41;
  C2 BY GRI5 GRI14 GRI27;
  C3 BY GRI1 GRI8 GRI13;
  C5 BY GRI7 GRI23;
  C6 BY GRI12 GRI32 GRI39;

  DEPEND BY C1 C2;
  LOC BY C3 C5 C6;
```

- But...our model ran into estimation issues for the factor covariance matrix – so we cannot use the results

```
THE MODEL ESTIMATION TERMINATED NORMALLY
```

```
WARNING:  THE LATENT VARIABLE COVARIANCE MATRIX (PSI) IS NOT POSITIVE
DEFINITE.  THIS COULD INDICATE A NEGATIVE VARIANCE/RESIDUAL VARIANCE FOR A
LATENT VARIABLE, A CORRELATION GREATER OR EQUAL TO ONE BETWEEN TWO LATENT
VARIABLES, OR A LINEAR DEPENDENCY AMONG MORE THAN TWO LATENT VARIABLES.
CHECK THE TECH4 OUTPUT FOR MORE INFORMATION.
PROBLEM INVOLVING VARIABLE C5.
```

# 1 Higher Order Factor: Mplus Syntax and Output

---

- And...more estimation difficulty this time – so we cannot compare the 1 higher order factor model to the saturated factor covariance model:

```
MODEL:
  C1 BY GRI26 GRI41;
  C2 BY GRI5 GRI14 GRI27;
  C3 BY GRI1 GRI13;
  C5 BY GRI7 GRI23;
  C6 BY GRI32 GRI39 GRI12;

  GAMBLE BY C1 C2 C3 C5 C6;
```

```
THE MODEL ESTIMATION TERMINATED NORMALLY
```

```
WARNING: THE LATENT VARIABLE COVARIANCE MATRIX (PSI) IS NOT POSITIVE
DEFINITE. THIS COULD INDICATE A NEGATIVE VARIANCE/RESIDUAL VARIANCE FOR A
LATENT VARIABLE, A CORRELATION GREATER OR EQUAL TO ONE BETWEEN TWO LATENT
VARIABLES, OR A LINEAR DEPENDENCY AMONG MORE THAN TWO LATENT VARIABLES.
CHECK THE TECH4 OUTPUT FOR MORE INFORMATION.
PROBLEM INVOLVING VARIABLE C5.
```

---

## **CONCLUDING REMARKS**

# Wrapping Up...

---

- Today was a discussion of multi-factor CFA
  - Most of the process hinges on having good single factor CFA models – the precursors to a multifactor CFA
- Important in analyses is to remember the steps
  - Create well-fitting separate single factor models
  - Combine into multifactor model
  - Modify model to make fit better
  - Determine a structural model for the factors is needed
    - ◆ Repeat the process – just with the factors
- Multifactor models are very difficult to construct
  - But can be very important