

---

# **Comparing Classical Test Theory with CFA -and- How To Use Test Scores in Secondary Analyses**

Latent Trait Measurement and  
Structural Equation Models

Lecture #8

March 6, 2013

# Today's Class

---

- Comparing classical test theory to CFA
  - The use and misuse of sum scores
  - Reliability for sum scores under CFA
- How to use CFA to test assumptions in CTT
- What to do when SEM isn't an option
  - Secondary analyses

# 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 scale of gambling tendencies using the first 24 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)

---

# CLASSICAL TEST THEORY

# *Classical* Test Theory (CTT)

---

- What you have learned about measurement so far likely falls under the category of CTT:
  - Writing items and building scales
  - Item analysis
  - Score interpretation
  - Evaluating reliability and construct validity
- Big picture: We will view CTT as model with a restrictive set of assumptions within a more general family of **latent trait measurement models**
  - Confirmatory Factor Analysis is a measurement model

# Differences Among Measurement Models

---

- What is the **name of the latent trait** measured by a test?
  - Classical Test Theory (CTT) = “True Score” (T)
  - Confirmatory Factor Analysis (CFA) = “Factor Score” (F)
  - Item Response Theory (IRT) = “Theta” ( $\theta$ )
- Fundamental difference in approach:
  - **CTT → unit of analysis is the WHOLE TEST** (item sum or mean)
    - ♦ **Sum = latent trait**, and the sum doesn’t care how it was created
    - ♦ Only using the sum requires restrictive assumptions about the items
  - **CFA, IRT, and beyond → unit of analysis is the ITEM**
    - ♦ Model of how item response relates to an **estimated latent trait**
    - ♦ Different models for differing item response formats
    - ♦ Provides a framework for testing adequacy of measurement models

# Classical Test Theory (CTT)

---

- In CTT, the **TEST** is the unit of analysis:  $Y_{total} = T + e$

- **True score T:**

- ♦ Best estimate of 'latent trait': Mean over infinite replications

- **Error e:**

- ♦ Expected value (mean) of 0, expected to be uncorrelated with T
- ♦ e's are supposed to wash out over repeated observations

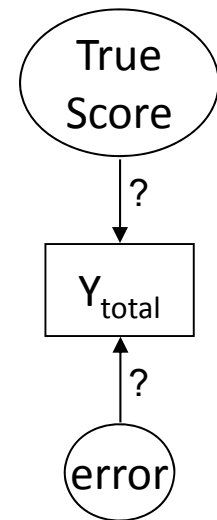
- **So the expected value of T is  $Y_{total}$**

- In terms of observed variance of the test scores:

- ♦ Observed variance = true variance + error variance

- Goal is to quantify ***reliability***

- ♦ Reliability = true variance / (true variance + error variance)



- Because the CTT model does not include individual items, **items must be assumed exchangeable** (and more items is better)

# Classical Test Theory, continued

---

- CTT unit of analysis is the WHOLE TEST (sum of items)
  - Want to ascertain how much of observed test score variance is due to 'true score' variance versus 'error' variance
  - Quantify 'error variance' in various ways
  - 'Error' is a unitary construct in CTT (and error is 'bad')
  - Goal is then to reduce 'error' variance as much as possible
    - ◆ Standardization of testing conditions (make confounds constants)
    - ◆ Aggregation = more items are better (errors should cancel out)
  - Items are exchangeable; properties are not taken into account
- Followed by *generalizability theory* to decompose error
  - e.g., rater variance, person variance, time variance

# Advantages of CFA over CTT

---

- More reasonable assumptions about items
  - CTT assumes tau-equivalent items
    - ◆ Tau – equivalent items: equal factor loadings
  - CFA allows a test of whether each item relates to the factor, as well as whether different factor loadings across items are needed
    - ◆ Would indicate some items are better than others
- Comparability across samples, groups, and time
  - CTT: No separation of observed item responses from true score
    - ◆ Sum across items = true score; item properties are for that sample only
  - CFA: Latent trait is estimated separately from item responses
    - ◆ Separates person traits from specific items given
    - ◆ Separates item properties from specific persons in sample
- Advantages apply to any latent trait model

# Reliability Measured by Alpha

---

- For quantitative items (items with a scale – although used on categorical items), this is Cronbach's Alpha...
  - Or 'Guttman-Cronbach alpha' (Guttman 1945 > Cronbach 1951)
  - Another reduced form of alpha for binary items: KR 20
- Alpha is described in multiple ways:
  - Is the mean of all possible split-half correlations
  - Is expected correlation with hypothetical alternative form of the same length
  - Is lower-bound estimate of reliability under assumption that all items are tau-equivalent (more about that later)
  - As an index of "internal consistency"
    - ♦ Although nothing about the index indicates consistency!

## Where Alpha Comes From

---

- The **sum of the item variances** is given by:
  - $\text{Var}(I_1) + \text{Var}(I_2) + \text{Var}(I_3) \dots + \text{Var}(I_k)$  (just the item variances)
- The **variance of the sum of the items** is given by the sum of ALL the item variances and covariances:
  - $\text{Var}(I_1 + I_2 + I_3) = \text{Var}(I_1) + \text{Var}(I_2) + \text{Var}(I_3) \dots$   
 $+ 2\text{Cov}(I_1, I_2) + 2\text{Cov}(I_1, I_3) + 2\text{Cov}(I_2, I_3) \dots$
  - Where does the '2' come from?
    - ♦ Covariance matrix is symmetric
    - ♦ Sum the whole thing to get to the variance of the sum of the items

	$I_1$	$I_2$	$I_3$
$I_1$	$\sigma_1^2$	$\sigma_{12}$	$\sigma_{13}$
$I_2$	$\sigma_{21}$	$\sigma_2^2$	$\sigma_{23}$
$I_3$	$\sigma_{31}$	$\sigma_{32}$	$\sigma_3^2$

# Guttman-Cronbach Alpha for Reliability

---

Covariance

Version:

k = # items

$$\alpha = \frac{k}{k-1} \cdot \frac{\text{variance of total } Y - \text{sum of item variances}}{\text{variance of total } Y}$$

- Numerator reduces to just the covariance among items
  - ***Sum of the item variances...***
    - ♦  $\text{Var}(X) + \text{Var}(Y) = \text{Var}(X) + \text{Var}(Y) \rightarrow$  just the item variances
  - ***Variance of total Y (the sum of the items)...***
    - ♦  $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y) \rightarrow$  **PLUS covariances**
  - So, if the items are related to each other, the variance of the total Y item sum should be bigger than the sum of the item variances
    - ♦ How much bigger depends on how much covariance among the items – the primary index of relationship

## Assessing Reliability in Our 12 Item Gambling Scale

---

- To get the Guttman-Chronbach Alpha of our 12 item scale, we need the covariance matrix
  - This can be found by the SAMPSTAT option under the OUTPUT statement
  - Sum of item variances = 11.834
  - Sum of item covariances = 21.575
  - Variance of Total Y =  $11.834 + 2 * 21.575 = 54.984$
- Alpha reliability:  $\rho_{\alpha} = \left( \frac{12}{12-1} \right) \left( \frac{54.984 - 11.834}{54.984} \right) = .852$

# Reliability...for what?

---

- The alpha reliability is the reliability for:
  - The total test score
  - Under the assumption that the items are **tau-equivalent**
    - ◆ Tau-equivalent means each item contributes equally
    - ◆ In a few slides, we will see how this translates to CFA
- What alpha is not:
  - An index of model fit (unidimensionality)

# Measurement Language: Don't Say These

---

- Often, people refer to items as “tapping” some latent trait
  - I think this makes the process less transparent – items measure the trait



- When alpha is used, you can sometimes hear people say something about how well the items “hang together”
  - This is certainly not true...

# How to Get Alpha UP

---

**TABLE 1**  
**Values of Cronbach's Alpha for Various Combinations of Different**  
**Number of Items and Different Average Interitem Correlations**

Number of Items	Average Interitem Correlation					
	.0	.2	.4	.6	.8	1.0
2	.000	.333	.572	.750	.889	1.000
4	.000	.500	.727	.857	.941	1.000
6	.000	.600	.800	.900	.960	1.000
8	.000	.666	.842	.924	.970	1.000
10	.000	.714	.870	.938	.976	1.000

# Alpha as Reliability... What could go wrong?

- Alpha does not index dimensionality → it does not index the extent to which items measure the same construct

**TABLE 13.2. Interitem Correlation Matrices for Two Hypothetical Tests with the Same Coefficient Alpha Reliability of .81**

Test A with 10 items											Test B with 6 items						
Variable	1	2	3	4	5	6	7	8	9	10	Variable	1	2	3	4	5	6
1.	—										1.	—					
2.	.3	—									2.	.6	—				
3.	.3	.3	—								3.	.6	.6	—			
4.	.3	.3	.3	—							4.	.3	.3	.3	—		
5.	.3	.3	.3	.3	—						5.	.3	.3	.3	.6	—	
6.	.3	.3	.3	.3	.3	—					6.	.3	.3	.3	.6	.6	—
7.	.3	.3	.3	.3	.3	.3	—										
8.	.3	.3	.3	.3	.3	.3	.3	—									
9.	.3	.3	.3	.3	.3	.3	.3	.3	—								
10.	.3	.3	.3	.3	.3	.3	.3	.3	.3	—							

- The variability
- We use inter

## Case In Point: All 24 Items

---

- Last class we showed that the 24 items of the GRI did not fit a one-factor model – what would happen if we neglected to check model fit and used the total score as our estimate of gambling tendency?
- The reliability estimate – from the covariance matrix of all the items (the saturated model H1) was  $\rho_{\alpha} = .861$ 
  - We would have concluded we had a “good” scale for gambling
- But, from CFA last week, we found that one factor didn’t describe all the items
  - Any subsequent analysis will have the misfit bias the results

# Testing CTT Assumptions in CFA

---

- Alpha is reliability assuming two things:
  - All factor loadings (discriminations) are equal, or that the items are “true-score equivalent” or “tau-equivalent”
  - Local independence (dimensionality now tested within factor models)
- We can test the assumption of tau-equivalence too via nested model comparisons in which the loadings are constrained to be equal – does model fit get worse?
  - If so, don’t use alpha – use model-based reliability (omega) instead. Omega assumes unidimensionality, but not tau-equivalence
  - Research has shown alpha can be an over-estimate or an under-estimate depending on particular data characteristics
- The assumption of ‘Parallel items’ is then testable by constraining item error variances to be equal, too – does model fit get worse?
  - ‘Parallel items’ will hardly ever hold in real data
  - Note that if tau-equivalence doesn’t hold, then neither does ‘parallel’

# Another Blast from the Past: Parallel Items

---

- Another CTT “model” that exists that of parallel items
  - All items have the same covariance and variance
  - Goes one step further than tau equivalence (equal covariances but unequal variances)
- Under the parallel items model, the alpha reliability for the total test score is called the Spearman-Brown reliability
  - Used to “prophesy” the number of items needed to increase reliability to a desired level
- Spearman-Brown Prophecy Formula
  - $\text{Reliability}_{\text{NEW}} = \text{ratio} * \text{rel}_{\text{old}} / [(\text{ratio}-1) * \text{rel}_{\text{old}} + 1]$ 
    - ♦ Ratio = ratio of new #items to old #items
  - For example:
    - ♦ Old reliability = .40
    - ♦ Ratio = 5 times as many items (had 10, what if we had 50)
    - ♦ New reliability = .77

# Reliability vs. Validity “Paradox”

---

- Given the assumptions of CTT, it can be shown that the correlation between a test and an outside criterion cannot exceed the reliability of the test (see Lord & Novick 1968)
  - Reliability of .81? No observed correlations possible  $> .9$ , because that's all the 'true' variance there to be relatable!
  - In practice, this may be false because it assumes that the errors are uncorrelated with the criterion (and they could be)
- Selecting items with the strongest discriminations (or the strongest inter-correlations) can help to 'purify' or homogenize a test, but potentially at the expense of construct validity
  - Can end up with a 'bloated specific'
  - Items that are least inter-related may be most useful in keeping the construct well-defined and thus relatable to other things

# Using CTT Reliability Coefficients: Back to the Score Estimates

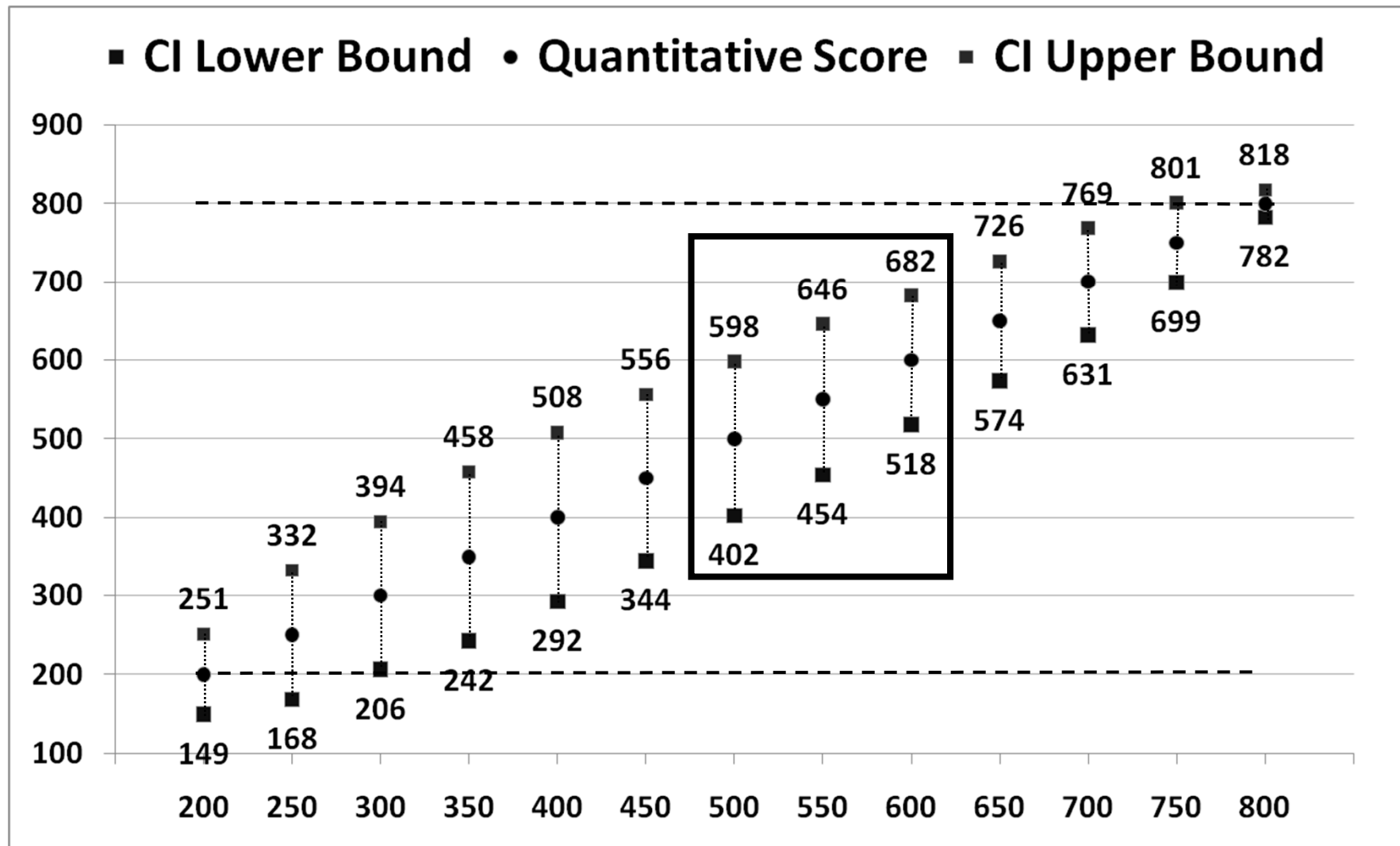
---

- Reliability coefficients are useful for describing the behavior of the test in the overall sample...  $\text{Var}(Y) = \text{Var}(T) + \text{Var}(e)$
- But reliability is a means to an end in interpreting a score for a given individual – we use it to get the error variance
  - $\text{Var}(T) = \text{Var}(Y) * \text{reliability}$ ; so  $\text{Var}(e) = \text{Var}(Y) - \text{Var}(T)$
  - **95% CI for individual score =  $Y \pm 1.96 * \text{SD}(e)$**
  - Gives an indication of how precise the true score estimate is on the metric of the original variable
  - Example:  $Y = 100, \text{Var}(e) = 9 \rightarrow 95\% \text{ CI} \approx 94 \text{ to } 106$   
 $Y = 100, \text{Var}(e) = 25 \rightarrow 95\% \text{ CI} \approx 90 \text{ to } 110$
  - Note this assumes a symmetric distribution, and thus will go out of bounds of the scale for extreme scores
  - Note this assumes the  $\text{SD}(e)$  or the **SE for each person is the same**
  - *Cue mind-blowing GRE example*

## 95% Confidence Intervals: Quantitative

*SEM ranges from 9 to 55*

---



---

# **REVISITING CTT FROM A CFA PERSPECTIVE**

## Classical Test Theory from a CFA Perspective

---

- In CTT the unit of analysis is the test score:

$$Y_{s,Total} = T_s + E_s$$

- In CFA the unit of analysis is the item:

$$Y_{si} = \mu_{I_i} + \lambda_i F_s + e_{si}$$

- To map CFA onto CTT, we must put these together:

$$Y_{s,Total} = \sum_{i=1}^I Y_{si}$$

## Further Unpacking of the Total Score Formula

---

- Because CFA is an item-based model, we can then substitute each item's model into the sum:

$$\begin{aligned} Y_{S,Total} &= \sum_{i=1}^I Y_{si} = \sum_{i=1}^I (\mu_{I_i} + \lambda_i F_S + e_{si}) \\ &= \sum_{i=1}^I \mu_{I_i} + \left( \sum_{i=1}^I \lambda_i \right) F_S + \sum_{i=1}^I e_{si} \end{aligned}$$

- Mapping this onto true score and error from CTT:

$$T = \sum_{i=1}^I \mu_{I_i} + \left( \sum_{i=1}^I \lambda_i \right) F_S \text{ and } E = \sum_{i=1}^I e_{si}$$

## Familiar Terms

---

- The tau-equivalent model assumes:
  - All items measure the factor the same:  $\lambda_i = \lambda$
  - Each item has its own unique variance:  $e_{si} \sim N(0, \psi_i^2)$
- The parallel items models assumes:
  - All items measure the factor the same:  $\lambda_i = \lambda$
  - All items have the same unique variance:  $e_{si} \sim N(0, \psi^2)$
- As such, each of these models can be tested by using the CFA approach – each are nested within the full CFA model

## Tau-Equivalence: Model Implied Covariance Matrix

---

- The CFA model implies a very specific form for the covariance matrix of the observed items:

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}^T + \mathbf{\Psi}$$

- The variance of an item  $i$  was:  $\sigma_i^2 = \lambda_i^2 \sigma_F^2 + \psi_i^2$
- The covariance of a pair of items  $i$  and  $j$  was:  $\sigma_{ij} = \lambda_i \lambda_j \sigma_F^2$
- Under Tau-Equivalence, all loadings are the same, meaning:
  - The item variances can be different (because of  $\psi_i^2$ )
  - All item covariances are the same ( $\lambda^2 \sigma_F^2$ )
- This is called the compound symmetry heterogeneous model
  - We can actually achieve the same model without the factor

# Tau-Equivalence Model for 12 Item Gambling Scale

- The following two pieces of Mplus syntax result in the same equivalent model:
- Tau-Equivalence as a Factor Model:

```

MODEL:
!LABELS FOR FACTOR LOADINGS IN PARENTHESES (SAME LABEL = SAME ESTIMATE)
  GAMBLING by GRI1* (loading)
               GRI3 (loading)
               GRI5 (loading)
               GRI6 (loading)
               GRI9 (loading)
               GRI10 (loading)
               GRI11 (loading)
               GRI13 (loading)
               GRI14 (loading)
               GRI15 (loading)
               GRI21 (loading)
               GRI23 (loading);

!UNIQUE VARIANCE LABELS:
  GRI1 (U1); GRI3 (U3); GRI5 (U5); GRI6 (U6); GRI9 (U9);
  GRI10 (U10); GRI11 (U11); GRI13 (U13); GRI14 (U14); GRI15 (U15);
  GRI21 (U21); GRI23 (U23);

!FACTOR VARIANCE SET TO 1:
GAMBLING@1;
  
```

MODEL FIT INFORMATION	
Number of Free Parameters	25
Loglikelihood	
H0 Value	-19051.350
H0 Scaling Correction Factor for MLR	2.5172

- Tau-Equivalence as a Compound-Symmetry Heterogeneous Variances Model:

```

MODEL:
  GRI1 GRI3 GRI5-GRI6 GRI9-GRI11 GRI13-GRI15 GRI21 GRI23 WITH
  GRI1 GRI3 GRI5-GRI6 GRI9-GRI11 GRI13-GRI15 GRI21 GRI23 (cov);
  
```

MODEL FIT INFORMATION	
Number of Free Parameters	25
Loglikelihood	
H0 Value	-19051.350
H0 Scaling Correction Factor for MLR	2.5172

# Model Implied Covariance Matrix

---

- All covariances equal/all variances different

	Model Estimated Covariances/Correlations/Residual Correlations				
	GRI1	GRI3	GRI5	GRI6	GRI9
GRI1	1.016				
GRI3	0.319	0.861			
GRI5	0.319	0.319	0.835		
GRI6	0.319	0.319	0.319	0.623	
GRI9	0.319	0.319	0.319	0.319	0.532
GRI10	0.319	0.319	0.319	0.319	0.319
GRI11	0.319	0.319	0.319	0.319	0.319
GRI13	0.319	0.319	0.319	0.319	0.319
GRI14	0.319	0.319	0.319	0.319	0.319
GRI15	0.319	0.319	0.319	0.319	0.319
GRI21	0.319	0.319	0.319	0.319	0.319
GRI23	0.319	0.319	0.319	0.319	0.319

	Model Estimated Covariances/Correlations/Residual Correlations				
	GRI10	GRI11	GRI13	GRI14	GRI15
GRI10	0.694				
GRI11	0.319	1.176			
GRI13	0.319	0.319	0.810		
GRI14	0.319	0.319	0.319	2.161	
GRI15	0.319	0.319	0.319	0.319	1.538
GRI21	0.319	0.319	0.319	0.319	0.319
GRI23	0.319	0.319	0.319	0.319	0.319

	Model Estimated Covariances/Correlations/Residual Correlations	
	GRI21	GRI23
GRI21	0.698	
GRI23	0.319	0.848

# Testing for Tau Equivalence

---

- The Tau-Equivalence model (assumed when you sum items) can be tested against the full CFA model
  - The models are nested, so we can use a likelihood ratio test
- Log-likelihood from CFA model: -18,988.425; SCF = 2.4309
  - 36 parameters (12 item intercepts, 11 factor loadings, 1 factor variance, 12 unique variances)
- Log-likelihood from TE model: -19,051.350; SCF = 2.5172
  - 25 parameters (12 item intercepts, 1 factor loading, 12 unique variances)
- MLR Likelihood ratio test:  $\chi^2 = 56.315, p < .001$
- Therefore, we reject the tau-equivalent model in favor of the CFA model – this means the simple sum of the items is not sufficient
  - We should use the CFA model factor score instead of a sum score

# Parallel Items: Model Implied Covariance Matrix

---

- The CFA model implies a very specific form for the covariance matrix of the observed items:

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}^T + \mathbf{\Psi}$$

- The variance of an item  $i$  was:  $\sigma_i^2 = \lambda_i^2 \sigma_F^2 + \psi_i^2$
- The covariance of a pair of items  $i$  and  $j$  was:  $\sigma_{ij} = \lambda_i \lambda_j \sigma_F^2$
- Under Parallel Items, all loadings and unique variances are the same:
  - All item variances are the same ( $\lambda^2 \sigma_F^2 + \psi^2$ )
  - All item covariances are the same ( $\lambda^2 \sigma_F^2$ )
- This is called the compound symmetry model
  - We can actually achieve the same model without the factor
- Because parallel items are nested within tau-equivalent items, we do not have to test this model as we know it will not fit when compared to the CFA model

# Test Score Reliability Under the CFA Model

---

- Coefficient alpha gave reliability for the total test score under the Tau-Equivalent Items Model
  - We rejected that model in favor of the CFA model
  - Therefore, coefficient alpha will not be correct for our total test score (if we were to still sum up the items)
- The notions of test score reliability under the CFA model now involve the factor loadings
  - But still come back to classical notion of reliability being the proportion of variance due to true score:  $\rho = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}$

## Deriving Reliability For Sum Scores Under the CFA Model

---

- To show where total-score reliability under the CFA model comes from, recall our CFA-model for the total score:

$$Y_{S,Total} = \sum_{i=1}^I \mu_{I_i} + \left( \sum_{i=1}^I \lambda_i \right) F_S + \sum_{i=1}^I e_{si}$$

- Mapping this onto true score and error from CTT:

$$T = \sum_{i=1}^I \mu_{I_i} + \left( \sum_{i=1}^I \lambda_i \right) F_S \text{ and } E = \sum_{i=1}^I e_{si}$$

- We now must derive the variance for T and E

## True Score Variance Under the CFA Model

---

- The variance for the true score:

$$\sigma_T^2 = Var(T) = Var\left(\sum_{i=1}^I \mu_{I_i} + \left(\sum_{i=1}^I \lambda_i\right) F_S\right)$$
$$Var\left(\left(\sum_{i=1}^I \lambda_i\right) F_S\right) = \left(\sum_{i=1}^I \lambda_i\right)^2 \sigma_F^2$$

## Error Variance Under the CFA Model

---

- Because the CFA model allows for the estimation of error covariances (although you shouldn't do that), the error variance under the CFA model becomes:

$$\begin{aligned}\sigma_E^2 &= Var(E) = Var\left(\sum_{i=1}^I e_{si}\right) \\ &= \sum_{i=1}^I \psi_i^2 + \sum_{i=1}^{I-1} \sum_{j=i+1}^I 2cov(i, j)\end{aligned}$$

- When error covariances are not estimated, the last term is zero, leaving  $Var(E) = \sum_{i=1}^I \psi_i^2$

## Reliability for Total Score Under CFA

---

- The reliability of the total score from CFA, is then:

$$\rho_{\omega} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2} = \frac{\left(\sum_{i=1}^I \lambda_i\right)^2 \sigma_F^2}{\left(\sum_{i=1}^I \lambda_i\right)^2 \sigma_F^2 + \sum_{i=1}^I \psi_i^2 + \sum_{i=1}^{I-1} \sum_{j=i+1}^I 2cov(i, j)}$$

- This reliability coefficient is called coefficient Omega ( $\omega$ )
- If the tau-equivalent model does not hold  $\rho_{\omega}$  is the reliability of a total test score (sum score)
  - Typically is higher than Alpha
  - If unidimensional model holds, coefficients will be close

# Calculating Omega for Our Test

- We can use Mplus to calculate Omega for our test:

```
MODEL:
!LABELS FOR FACTOR LOADINGS IN PARENTHESES (SAME LABEL = SAME ESTIMATE)
  GAMBLING by GRI1
    GRI3 (L3)
    GRI5 (L5)
    GRI6 (L6)
    GRI9 (L9)
    GRI10 (L10)
    GRI11 (L11)
    GRI13 (L13)
    GRI14 (L14)
    GRI15 (L15)
    GRI21 (L21)
    GRI23 (L23);

!UNIQUE VARIANCE LABELS:
  GRI1 (U1); GRI3 (U3); GRI5 (U5); GRI6 (U6); GRI9 (U9);
  GRI10 (U10); GRI11 (U11); GRI13 (U13); GRI14 (U14); GRI15 (U15);
  GRI21 (U21); GRI23 (U23);

!FACTOR VARIANCE SET TO 1:
  GAMBLING (G_VAR);

MODEL CONSTRAINT:
  NEW (OMEGA);

OMEGA = ( G_VAR*(1+L3+L5+L6+L9+L10+L11+L13+L14+L15+L21+L23)^2) /
  ((G_VAR*(1+L3+L5+L6+L9+L10+L11+L13+L14+L15+L21+L23)^2 +
  (U1+U3+U5+U6+U9+U10+U11+U13+U14+U15+U21+U23)));
```

New/Additional Parameters				
OMEGA	0.855	0.009	91.688	0.000

Here, Omega is .855

# Omega Under Tau-Equivalent Items

- Omega equal to Alpha when you use the tau-equivalent items model
  - Omega is the Spearman Brown reliability under parallel items

```
MODEL:
!LABELS FOR FACTOR LOADINGS IN PARENTHESES (SAME LABEL = SAME ESTIMATE)
  GAMBLING by GRI1* (loading)
               GRI3 (loading)
               GRI5 (loading)
               GRI6 (loading)
               GRI9 (loading)
               GRI10 (loading)
               GRI11 (loading)
               GRI13 (loading)
               GRI14 (loading)
               GRI15 (loading)
               GRI21 (loading)
               GRI23 (loading);

!UNIQUE VARIANCE LABELS:
  GRI1 (U1); GRI3 (U3); GRI5 (U5); GRI6 (U6); GRI9 (U9);
  GRI10 (U10); GRI11 (U11); GRI13 (U13); GRI14 (U14); GRI15 (U15);
  GRI21 (U21); GRI23 (U23);

!FACTOR VARIANCE SET TO 1:
  GAMBLING@1;

MODEL CONSTRAINT:
  NEW (OMEGA);
  OMEGA = ( 1*(12*LOADING)^2)/
           ( (1*(12*LOADING)^2 +
             (U1+U3+U5+U6+U9+U10+U11+U13+U14+U15+U21+U23) ) );|

New/Additional Parameters
  OMEGA                0.852        0.011        80.427        0.000
```

Here, Omega is .852 – which is equal to the Alpha we calculated using the covariance matrix

## Recapping: CTT using CFA

---

- Classical test theory – and more specifically, total test scores, is the dominant way to assess subjects
  - This is true even under CFA
- The key is to be sure to check if a one-factor model fits the data before using any type of reliability coefficient
  - If not, do not use a test score
- If the one factor model fits – then a single score can represent the test
- The next worry is about representing the error in the test score (related to reliability)
  - If reliability is “high” (? How high, standard of .8), then using the test score in a subsequent analysis is accepted practice

## Secondary Analyses with Factor Scores

---

- If you want to use results from a survey in a new analysis
  - Best: Use SEM – error in factor scores is already partitioned variance
  - Similarly good: Use “plausible values” (repeated draws from posterior distribution of each person’s factor score) – essentially what SEM does – but with factor scores that vary within a person
    - ♦ Can be done in Mplus – not described
  - Slightly Less Good: Use SEM with “single indicator” factors using sum scores
    - ♦ The focus of the next section
    - ♦ Make error variance =  $(1 - \text{reliability}) * \text{Variance (Sum score)}$ ; factor loading = 1
  - Okay (but widespread): for scales that are unidimensional (and verified in CFA), use sum scores
    - ♦ Assumes unidimensionality and “high” reliability
  - Not Cool: Use factor scores only

# What about Using Factor Scores?

---

- Although CFA factor scores have fewer problems than EFA factor scores (because there is no rotation in CFA), they still have issues:
  - They will be **shrunk** (i.e., pushed towards the mean, such that the observed variance of the factor scores will be less than the original factor variance)
  - Can get estimates of “**factor determinacy**” → how correlated estimated factor scores are with true factor scores (basically how much error is introduced by estimating the factor scores as observed variables)
  - They are just **estimates of central tendency** from a distribution for each person, not known values – and using estimates as known values in another model makes the relationships within that model look more precise than they are (like SE = 0)
  - You CANNOT create factor scores by using the loadings as such:
    - ♦  $F = \lambda_{11}y_1 + \lambda_{21}y_2 + \lambda_{31}y_3 \dots$  → This is a COMPONENT model, not a FACTOR model.

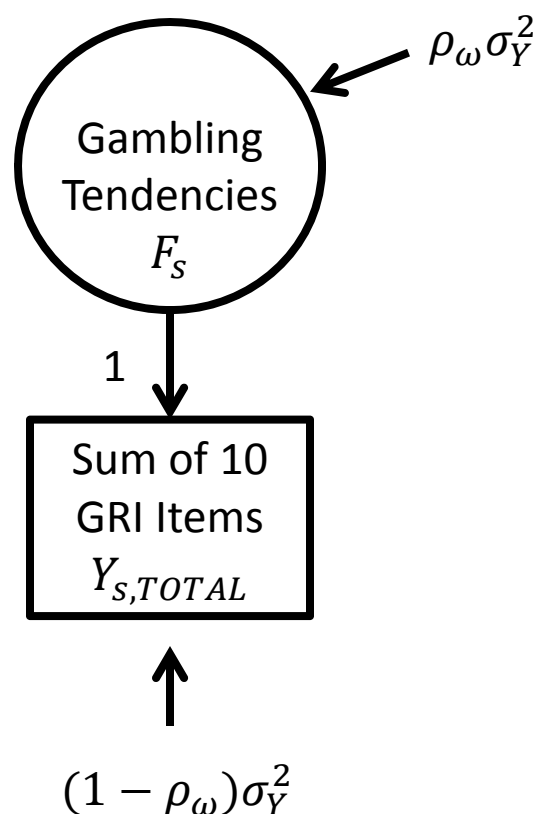
---

# **SINGLE INDICATOR MODELS**

# Single Indicator Models

---

- Single indicator models are CFA-like models where a “factor” is measured by a single indicator:
  - Shown here for the gambling factor



# Identification in Single Indicator Models

---

- How is this possible? Isn't a single indicator factor model unidentified?
  - We fix the factor variance, factor loading, and unique variance
  - Factor variance represents “reliable” portion
- Single indicator model parameters:
  - $\sigma_F^2$  - factor variance;  $\lambda_Y$  - factor loading;  $\psi_Y^2$  - item unique variance (assume factor mean fixed to zero and item intercept is set to its mean)
- Our constraints are:
  - $\lambda_Y = 1$
  - $\sigma_F^2 = \rho_\omega * \sigma_Y^2$  (the portion of Y that is “reliable”)
  - $\psi_Y^2 = (1 - \rho_\omega) * \sigma_Y^2$  (the portion of Y that is left over)

# Assumptions in a Single Indicator Analysis

---

- To use a single indicator you must assume:
  - The indicator is unidimensional (only one factor)
    - ♦ This is testable in CFA (but if you have a small sample is hard to do)
    - ♦ If not possible to test, you must assume you have one factor
      - This is an assumption that the test is \*as\* dimensional in your sample/population
  - The reliability of the indicator is known
    - ♦ Also obtainable from CFA
    - ♦ If not possible to obtain, then you must use a previously reported reliability coefficient
      - This is an assumption that the test is \*as\* reliable in your sample/population

# Single Indicator Example Analysis

---

- To demonstrate a single indicator example analysis, we will use the 12-item GRI to predict the SOGS score
  - SOGS = South Oak Gambling Screen (we collected this)
    - ◆ Note: we assume this has reliability of 1.0
  - The 12-item GRI is the single indicator of the gambling factor
- Step #1: determine that the single indicator is unidimensional
  - The 1-factor CFA model fit the 12-item GRI
- Step #2: get the single indicator reliability
  - From the CFA analysis we found that the reliability of the 12-item GRI was .855
- Step #3: estimate the variance of the 12-item GRI total score
  - We can do this in Mplus – found the variance to be 55.050

# Single Indicator Analysis

---

- Now that we have our reliability of the GRI and the variance of the GRI, we can put these into the single indicator model:

```
VARIABLE:
  NAMES = GRI1-GRI41 SOGS4-SOGS15 Student ID;
  USEVARIABLES = GRI12sum SOGSsum;
  IDVARIABLE = ID;
  MISSING = ALL(99);

DEFINE:
  SOGSsum = SUM(SOGS4-SOGS15);
  GRI12sum = SUM(GRI1 GRI3 GRI5-GRI6 GRI9-GRI11 GRI13-GRI15 GRI21 GRI23);

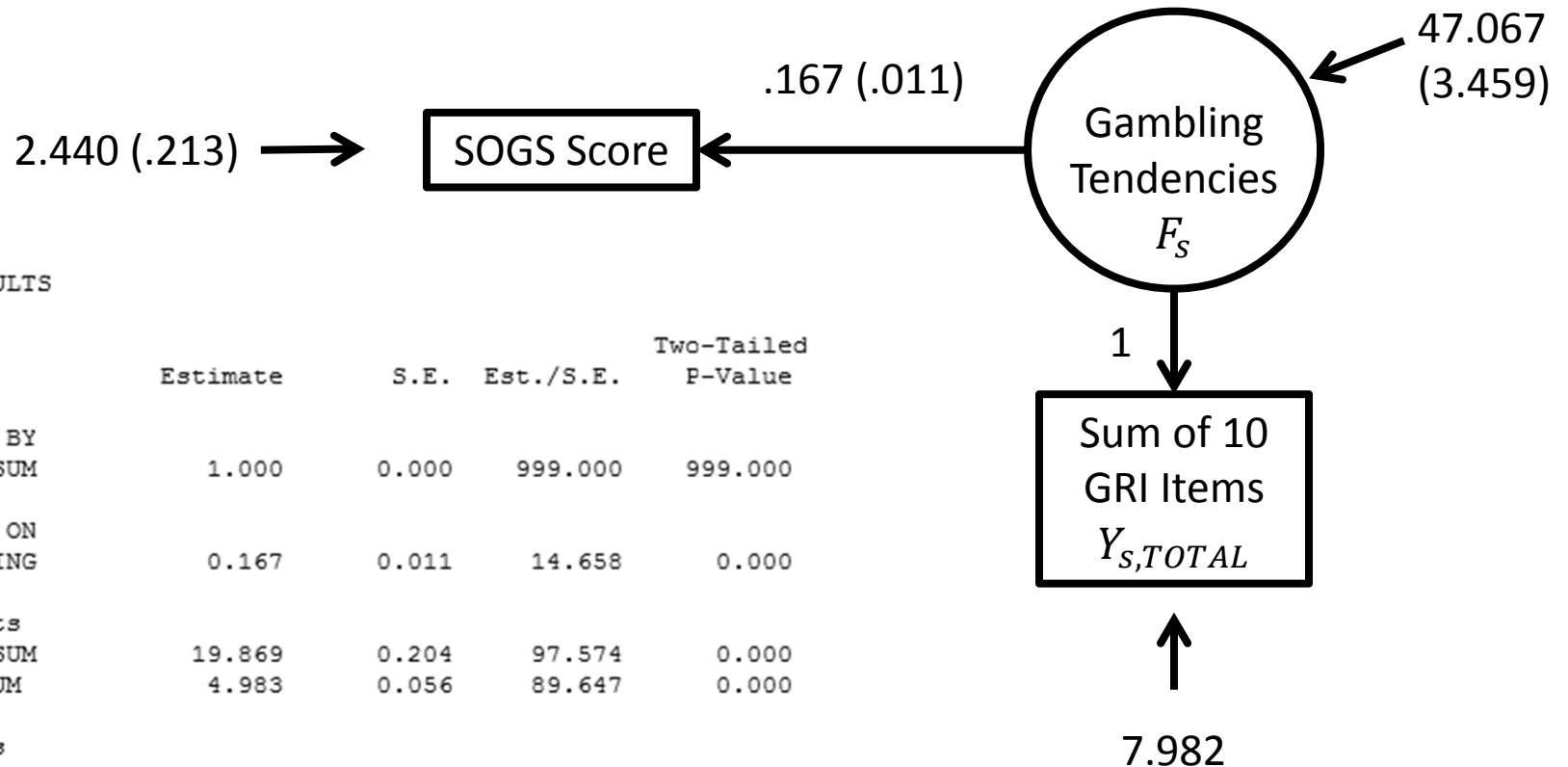
MODEL:
  !DEFINE GAMBLING FACTOR
  GAMBLING BY GRI12sum@1; !FIX FACTOR LOADING TO 1
  GRI12sum (si_var);      !LABEL UNIQUE VARIANCE TO CONSTRAIN IT LATER

  SOGSsum ON GAMBLING;    !MAKE PREDICTION OF SOGSSUM USING
                          !GAMBLING "LATENT" VARIABLE

MODEL CONSTRAINT:
  !HERE THE SINGLE INDICATOR VARIANCE IS SET
  !THIS IS THE VARIANCE OF THE SUM (54.858) TIMES 1-OMEGA RELIABILITY (.855)
  si_var = 55.050*(1-.855);
```

# Single Indicator Model Results:

- Note:  $47.067 + 7.982 = 55.05$  (the variance of the GRI)



## MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
<b>GAMBLING BY</b>				
GRI12SUM	1.000	0.000	999.000	999.000
<b>SOGSSUM ON</b>				
GAMBLING	0.167	0.011	14.658	0.000
<b>Intercepts</b>				
GRI12SUM	19.869	0.204	97.574	0.000
SOGSSUM	4.983	0.056	89.647	0.000
<b>Variances</b>				
GAMBLING	47.067	3.459	13.608	0.000
<b>Residual Variances</b>				
GRI12SUM	7.982	0.000	999.000	999.000
SOGSSUM	2.440	0.213	11.445	0.000

# Single Indicator Model Interpretation

- The standardized regression slope for the gambling “factor”, predicting the SOGS was .591 – as gambling went up 1 SD, the SOGS score went up .591 SD

- Correlation between gambling and SOGS

- The gambling “factor” accounted for 35.0% of the variance in the SOGS score

## STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GAMBLING BY GRI12SUM	0.925	0.005	187.698	0.000
SOGSSUM ON GAMBLING	0.591	0.035	17.010	0.000
Intercepts				
GRI12SUM	2.678	0.069	38.730	0.000
SOGSSUM	2.572	0.092	27.967	0.000
Variances				
GAMBLING	1.000	0.000	999.000	999.000
Residual Variances				
GRI12SUM	0.145	0.009	15.916	0.000
SOGSSUM	0.650	0.041	15.804	0.000

## R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GRI12SUM	0.855	0.009	93.849	0.000
SOGSSUM	0.350	0.041	8.505	0.000

# Comparison with Non-Single Indicator

- Without using the single indicator:

## MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value	STDYX Standardization		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SOGSSUM ON GRI12SUM	0.143	0.010	13.877	0.000						
Means					SOGSSUM ON GRI12SUM	0.547	0.034	16.173	0.000	
GRI12SUM	19.869	0.204	97.574	0.000	Means					
Intercepts					GRI12SUM	2.678	0.069	38.730	0.000	
SOGSSUM	2.145	0.183	11.733	0.000	Intercepts					
Variances					SOGSSUM	1.107	0.131	8.428	0.000	
GRI12SUM	55.050	3.459	15.916	0.000	Variances					
Residual Variances					GRI12SUM	1.000	0.000	999.000	999.000	
SOGSSUM	2.630	0.218	12.071	0.000	Residual Variances					
					SOGSSUM	0.701	0.037	18.946	0.000	

## R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SOGSSUM	0.299	0.037	8.086	0.000

---

# ITEM PARCELING

# Item Parceling

---

- Frequently, sum-scores are used in SEM under a different label: as item “parcels”
  - Evidently parcel sounds more polite “stuff that didn’t fit”
- Item parcels are sums of sets of items that are inserted into a SEM without any further inspection
- Frequently, item parcels will hide bad fit of model
  - Blind parceling = cheating
- As parcels are sums – and today’s class is about using sums, we can now discuss parcels under CTT with CFA

# Applying Our Understanding of Total/Sum Scores to Parcels

---

- As we have seen today, a total score is a statistical model
  - Tau-equivalent items
- As with any statistical model, if the model does not fit (adequately represent the data), misleading results occur
- Parceling items makes an implicit assumption about their structure – that they too are tau-equivalent
  - If that assumption is not valid, results cannot be believed
- Most uses of parceling make no attempt to determine if the tau-equivalence assumption is correct

## Revisiting our 24 Item GRI

- As you will recall, our 24 item GRI did not fit the one-factor model
- To demonstrate parceling, we will take the 12 misfitting items and create a parcel (sum their scores together)
  - We will then add the parcel to a CFA model with the other 12 items

### Chi-Square Test of Model Fit

Value	2697.493*
Degrees of Freedom	252
P-Value	0.0000
Scaling Correction Factor for MLR	1.3825

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	0.086
90 Percent C.I.	0.083 0.089
Probability RMSEA <= .05	0.000

The model RMSEA indicated the model did not fit well (want this to be < .05)

### CFI/TLI

CFI	0.684
TLI	0.654

The model CFI and TLI indicated the model did not fit well (want these to be > .95)

### SRMR (Standardized Root Mean Square Residual)

Value	0.087
-------	-------

The SRMR indicated the model did not fit well (want this to be < .08)

# The 12 Item GRI Plus the Parcel of Bad Items

---

- The syntax:

```
VARIABLE:
  NAMES = GRI1-GRI41 SOGS4-SOGS15 Student ID;
  USEVARIABLES = GRI1 GRI3 GRI5-GRI6 GRI9-GRI11 GRI13-GRI15 GRI21 GRI23
                GRIrest;
  IDVARIABLE = ID;
  MISSING = ALL(99);

DEFINE:
  GRIrest = SUM(GRI2 GRI4 GRI7 GRI8 GRI12 GRI16-GRI20 GRI22 GRI24);

MODEL:
  GAMBLING by GRI1 GRI3 GRI5-GRI6 GRI9-GRI11 GRI13-GRI15 GRI21 GRI23 GRIrest;
```

- The model fit statistics (adequate model fit):

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

      Estimate                0.055
      90 Percent C.I.         0.049  0.061
      Probability RMSEA <= .05 0.082

CFI/TLI

      CFI                0.929
      TLI                0.915
```

- Note: 12 item GRI had RMSEA of .045

## How Parceling Hides Poor Model Fit

---

- The item parcel hides poor model fit by using a numbers game to its advantage
- Model with 24 items had 300 elements in saturated covariance matrix (but 48 parameters for that matrix)
- Model with 12 items plus parcel (12 “items”) had 91 elements in saturated covariance matrix (and 26 parameters)
- The relative ratio of parameters to saturated covariances makes the parcel hide the fit issues
  - Especially when the remainder of the items fit well already

# Parceling Done Right

---

- To add a parcel you must first examine the fit of a one-factor model to the items of the parcel:

```
VARIABLE:
  NAMES = GRI1-GRI41 SOGS4-SOGS15 Student ID;
  USEVARIABLES = GRI2 GRI4 GRI7 GRI8 GRI12 GRI16-GRI20 GRI22 GRI24;
  IDVARIABLE = ID;
  MISSING = ALL(99);

MODEL:
  PARCELI by GRI2 GRI4 GRI7 GRI8 GRI12 GRI16-GRI20 GRI22 GRI24;
```

- The model fit suggests a one-factor model doesn't fit

## Chi-Square Test of Model Fit

Value	1831.999*
Degrees of Freedom	54
P-Value	0.0000
Scaling Correction Factor for MLR	1.0632

## RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.157
90 Percent C.I.	0.151 0.163
Probability RMSEA <= .05	0.000

## CFI/TLI

CFI	0.357
TLI	0.214

# Why Parceling is Cheating and Why You Shouldn't Do It

---

- If you didn't check the parcel before adding it to the 12 item GRI model you would conclude 1-factor model fit the data well
  - If a one-factor model fits, then what comes next is typically the use of its sum score
- The sum score from the 12 good + 1 bad (parcel) model is just the sum score from the 24 item GRI – which didn't fit a one-factor model
  - The caveat: the 12+1 model sum score had an omega reliability of .516!
  - Most of the lack of reliability comes from the estimated unique variance of the parcel
- You cannot make a good factor by cheating with a parcel!

---

## **CONCLUDING REMARKS**

# Wrapping Up

---

- Today was spent on comparing classical test theory (synonymous with sum scores) to CFA
- Understanding how CTT and CFA are related is important
  - Many people believe that sum scores are A-OK
    - ♦ They only are if they fit a 1-factor model and have a high reliability
  - Many people don't think parceling involves sum scores
    - ♦ The label must be the problem...
- Single indicator models can be a good way to use sum scores if:
  - The 1-factor model fits
  - There is a high degree of reliability
  - We will return to this once we discuss SEM more thoroughly

## Coming Up...

---

- Next week's lecture: multidimensional CFA models
  - More than one factor
  - Reliability for a total test score no longer applies (each factor is where reliability is important)
  - Time permitting: an introduction to Exploratory Factor Analysis
    - ◆ Followed by a comparison of CFA and EFA
    - ◆ And why you also shouldn't be doing EFA
      - But could explore the data better using CFA techniques