Measurement Invariance (MI) in CFA and Differential Item Functioning (DIF) in IRT/IFA

- Topics:
  - What are MI and DIF?
  - Testing measurement invariance in CFA
  - Testing differential item functioning in IRT/IFA

The Big Picture

- **In CFA**, we are assessing "**measurement invariance**" (**MI**), also known as "factorial invariance" or "measurement equivalence"

- Concerns the extent to which are the psychometric properties of the observed indicators are transportable or generalizable across groups (e.g., gender, language) or over time/conditions
  - In other words, we are testing whether the indicators measure the same construct *in the same way* in different groups or over time/condition
  - If so, then indicator responses should depend only on latent trait scores, and not on group membership or time/condition, such that observed response differences are caused by TRUE differences in the trait

- **In IRT/IFA**, lack of measurement invariance is known as "**differential item functioning**" (**DIF**), but it's the same idea
  - Note the inversion: Measurement Invariance = Non-DIF
    Measurement Non-Invariance = DIF

2 Distinct Types of Invariance

- **Measurement Invariance** concerns how the indicators measure the latent trait across groups or time/condition
  - An invariant measurement model has the same factor **loadings**, item **intercepts/thresholds**, and **residual variances** (and covariances)
  - Measurement model invariance is a precursor to ANY group or time/condition comparison (whether explicitly tested or not)
  - **It's not ok** if you don't have at least partial measurement invariance to make subsequent comparisons across groups or time/condition

- **Structural Invariance** concerns how the latent traits are distributed and related across groups or time/condition
  - An invariant structural model has the same **factor variances, factor covariances** (or same higher-order structure) and **factor means**
  - Given (at least partial) measurement invariance, **it is ok** if you don't have structural invariance, **because those trait differences may be real**

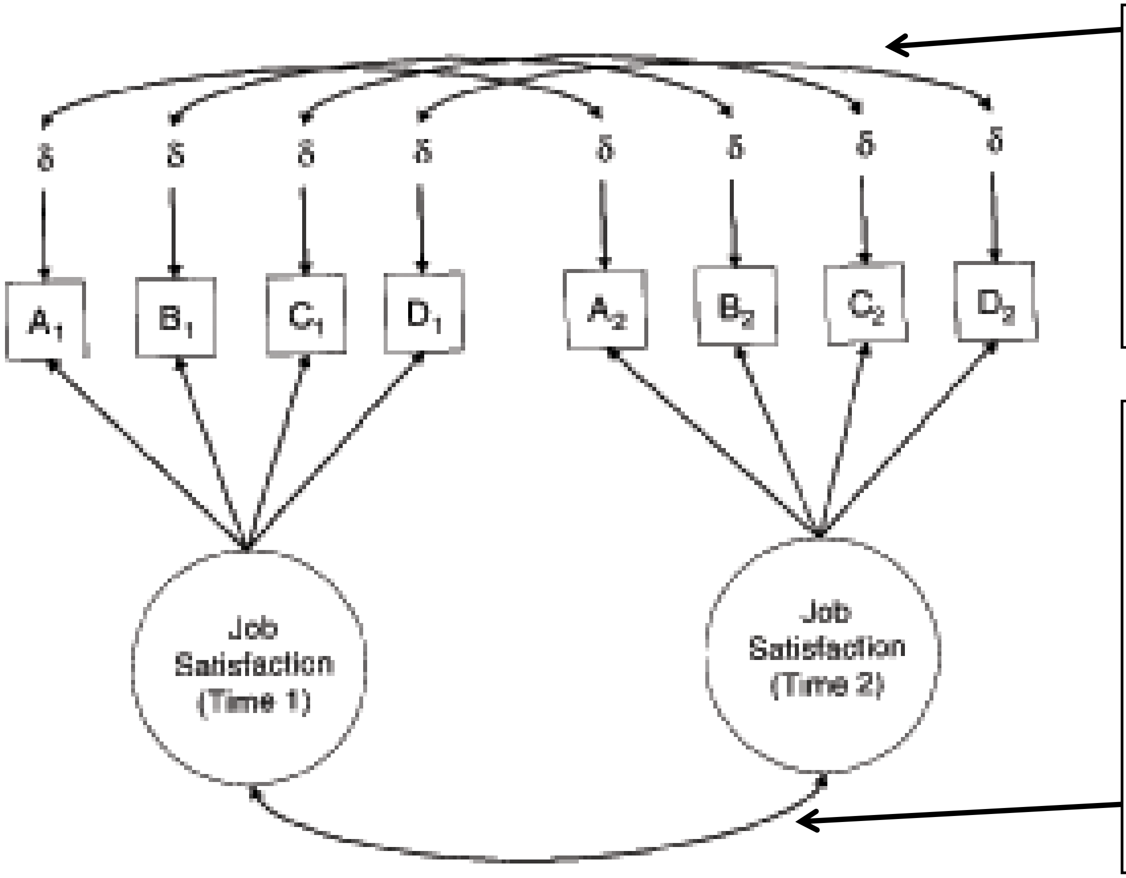
Model Options for Testing Invariance

- Invariance testing in CFA (or testing DIF in IRT/IFA) proceeds differently depending on the type of groups to be compared

- **Independent groups?** Use a "**multiple-group**" model
  - Test separate group-specific factor models, but **simultaneously**
  - Use GROUP = in Mplus and separate MODEL statements per group
  - An alternative approach, MIMIC models, in which the grouping variable is entered as a predictor, do not allow testing of equality of factor loadings or factor variances (so MIMIC is less useful than a full multiple-group model)

- **Dependent** (longitudinal, repeated, dyadic) groups?
  - All indicator responses go into **SAME model**, with separate factors per occasion/condition (allowing all factor covariances by default)
  - Usually, same indicators also have residual covariances by default
  - Given measurement invariance, growth modeling of the latent traits can serve as a specific type of structural invariance testing
  - It is INCORRECT to use a multiple-group model if groups are dependent

Longitudinal Invariance Model



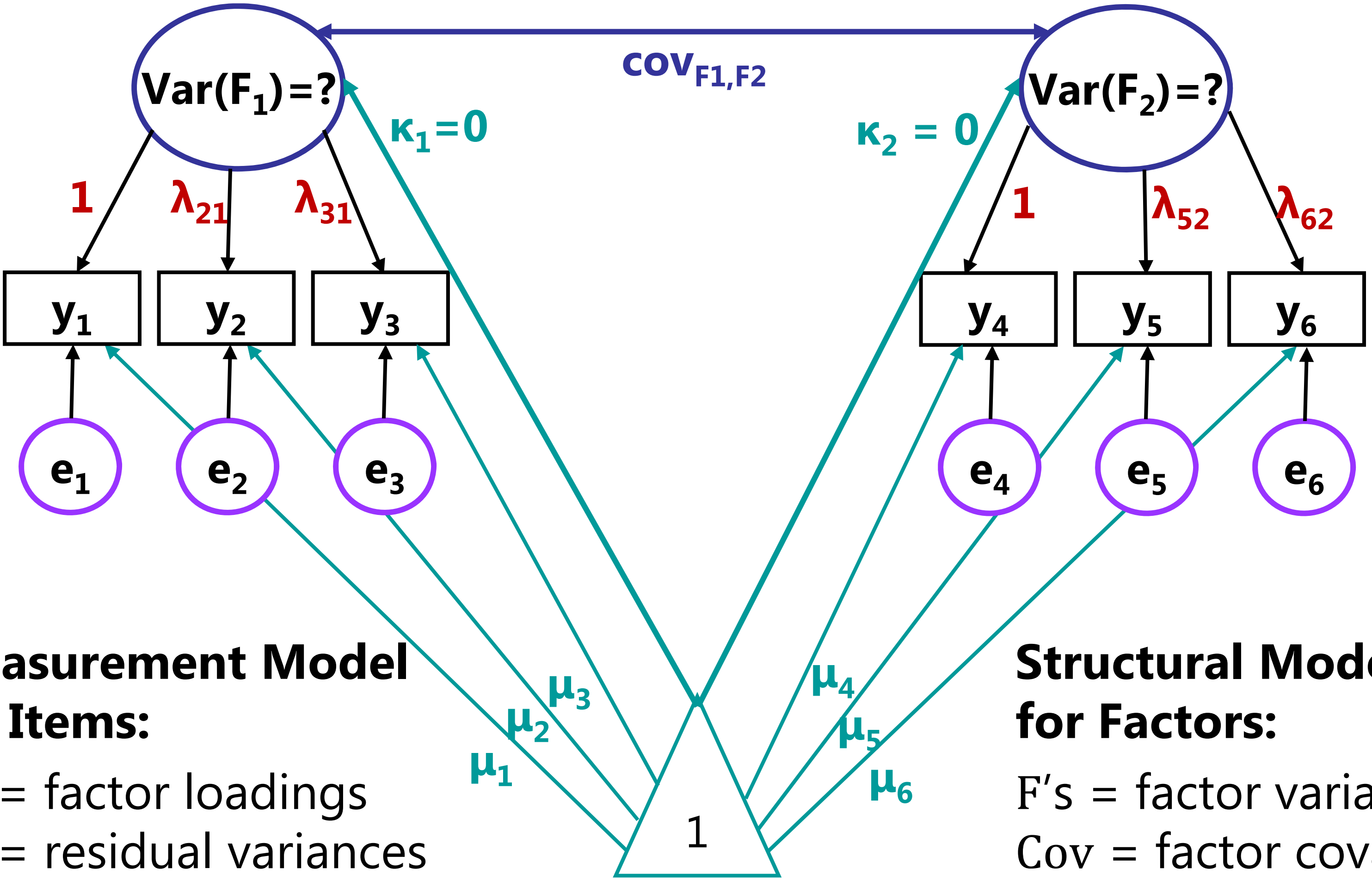
Residual covariances for same indicators at different repeated measurements are often included by default

Factors are estimated separately for each repeated measurement and covariances are always estimated to reflect dependency of observations

FYI: A structural model in which all factor means, variances, and covariances are estimated is analogous to a "saturated means, unstructured variance model" for observed variables in MLM terms

Remember the CFA model? Let's start MI testing here....

We will begin with the Mplus default of a marker item loading but a 0 factor mean.



Measurement Model for Items:

λ's = factor loadings
e's = residual variances
μ's = intercepts

Structural Model for Factors:

F's = factor variances
Cov = factor covariances
K's = factor means

Steps of Testing Invariance across Groups

- **Step 0: Omnibus test of equality of the overall indicator covariance matrix across groups**
  - Do the covariances matrices differ between groups, on the whole?
  - If not, game over. You are done. You have invariance. Congratulations.
  - Many people disagree with the necessity or usefulness of this test to begin testing invariance… why might that be?
  - People also differ in whether invariance should go from top-down or bottom-up directions… I favor bottom-up for the same reason.

- Let's proceed with an example with 2 factors, 6 indicators (3 per factor), and 2 groups…
  - Total possible # parameters = $\frac{v(v+1)}{2} + v = \frac{6(6+1)}{2} + 6 = 27$ per group
  - So our COMBINED possible DF = 54 across 2 groups

Step 1: Test “Configural” Invariance

- Do the groups have the same general factor structure?
- Same number of factors, same pattern of free/0 loadings → same conceptual definition of latent constructs
- In practice, begin by testing factor structure within each group separately, hoping they are “close enough”
- Then estimate separate group-specific models simultaneously, but **allow all model parameters to differ across groups**
  - This will be the baseline model for further comparisons
  - χ2 and df will be additive across groups (different group sample sizes will result in differential weighting of χ2 across groups)
- This is as good fit as it gets! From here forward, our goal is to make model fit NOT WORSE by constraining parameters equal
  - That means if the configural model fits badly, game over…

Testing Invariance Constraints

- As before, we will test whether subtracting parameters worsens model fit via likelihood ratio (aka, $-2\Delta LL$, χ^2) tests
  - Implemented via a direct difference in H0 model χ^2 values most often, but this is only appropriate when using regular ML estimation

- MLR requires a modified version of this $-2\Delta LL$ test (see Mplus website): http://www.statmodel.com/chidiff.shtml
  - Is called "rescaled likelihood ratio test" when you explain it
  - Includes extra steps to incorporate scaling factors (1.00 = regular ML)
  - I built you a spreadsheet for this…you're still welcome ☺

- If **removing parameters** (e.g., in invariance testing), H0 model fit can get **worse OR not worse** (as indicated by smaller LL OR by larger $-2LL$ and χ^2)
- If **adding parameters** (e.g., in adding factors), H0 model fit can get **better OR not better** (as indicated by larger LL OR by smaller $-2LL$ and χ^2)

Testing Fixes to the Model: −2ΔLL

- Comparing nested models via a "**likelihood ratio test**" → **−2ΔLL** (MLR rescaled version)
  - 1. Calculate **−2ΔLL** = $-2*(LL_{fewer} - LL_{more})$
  - 2. Calculate **difference scaling correction** =

$$\frac{(\#parms_{fewer}*scale_{fewer}) - (\#parms_{more}*scale_{more})}{(\#parms_{fewer} - \#parms_{more})}$$

  - 3. Calculate **rescaled difference** = −2ΔLL / scaling correction
  - 4. Calculate **Δdf** = $\#parms_{more} - \#parms_{fewer}$
  - 5. **Compare rescaled difference to χ^2 with df = Δdf**
    - Add 1 parameter? $LL_{diff} > 3.84$, add 2 parameters: $LL_{diff} > 5.99$…
    - Absolute values of LL are meaningless (is relative fit only)
    - Process generalizes to many other kinds of models

Note: Your LL will always be listed as the **H0**
(H1 is for the saturated, perfectly fitting model)

Fewer = simpler model
More = more parameters

1. **Configural** Invariance Model:

Same Factor Structure; All Parameters Separate

Total DF across groups = 54 – 38 = 16 =

$$54-\left(12\mu+12\sigma_e^2+8\lambda+4\sigma_F^2+2\sigma_{F12}+0\kappa_F\right)=16$$

Group 1 (subscript = item, group):

- $y_{11} = \mu_{11} + \mathbf{1}F_1 + e_{11}$
- $y_{21} = \mu_{21} + \lambda_{21}F_1 + e_{21}$
- $y_{31} = \mu_{31} + \lambda_{31}F_1 + e_{31}$
- $y_{41} = \mu_{41} + \mathbf{1}F_2 + e_{41}$
- $y_{51} = \mu_{51} + \lambda_{51}F_2 + e_{51}$
- $y_{61} = \mu_{61} + \lambda_{61}F_2 + e_{61}$
- Both factors have estimated variances and a covariance, but both factor means are fixed to 0

Group 2 (subscript = item, group):

- $y_{12} = \mu_{12} + \mathbf{1}F_1 + e_{12}$
- $y_{22} = \mu_{22} + \lambda_{22}F_1 + e_{22}$
- $y_{32} = \mu_{32} + \lambda_{32}F_1 + e_{32}$
- $y_{42} = \mu_{42} + \mathbf{1}F_2 + e_{42}$
- $y_{52} = \mu_{52} + \lambda_{52}F_2 + e_{52}$
- $y_{62} = \mu_{62} + \lambda_{62}F_2 + e_{62}$
- Both factors have estimated variances and a covariance, but both factor means are fixed to 0

Step 2: Test "**Metric**" Invariance

- Also called "**weak** factorial invariance"
- Do the groups have the same **factor loadings**?
  - Each "congeneric" indicator is still allowed to have a different loading (i.e., this is not a tau-equivalent model)
  - Loadings for same indicator are constrained equal across groups

- Change the method of model identification with respect to the factor loadings and factor variances only: Estimate all newly constrained factor loadings, but **fix the factor variances to 1 in the reference group** (free factor variances in other group)
  - Why? Loadings for marker items (fixed=1 for identification) would be assumed invariant, and thus they could not be tested
  - This alternative specification allows us to evaluate ALL loadings and still identify the model (see Yoon & Millsap, 2007), which is BETTER

2. **Metric** Invariance Model:

Same Factor Loadings Only (saves 4 df)

Total DF across groups = 54 – 34 = 20 $=$

$$54 - \left(12\mu + 12\sigma^2_e + \mathbf{6\lambda} + \mathbf{2\sigma^2_F} + 2\sigma_{F12} + 0\kappa_F\right) = 20$$

<u>Group 1 (subscript = item, group):</u>

- $y_{11} = \mu_{11} + \boldsymbol{\lambda_1 F_1} + e_{11}$
- $y_{21} = \mu_{21} + \boldsymbol{\lambda_2 F_1} + e_{21}$
- $y_{31} = \mu_{31} + \boldsymbol{\lambda_3 F_1} + e_{31}$
- $y_{41} = \mu_{41} + \boldsymbol{\lambda_4 F_2} + e_{41}$
- $y_{51} = \mu_{51} + \boldsymbol{\lambda_5 F_2} + e_{51}$
- $y_{61} = \mu_{61} + \boldsymbol{\lambda_6 F_2} + e_{61}$
- **Both factor variances fixed to 1 for identification**, factor covariance is estimated, but both factor means are STILL fixed to 0

<u>Group 2 (subscript = item, group):</u>

- $y_{12} = \mu_{12} + \boldsymbol{\lambda_1 F_1} + e_{12}$
- $y_{22} = \mu_{22} + \boldsymbol{\lambda_2 F_1} + e_{22}$
- $y_{32} = \mu_{32} + \boldsymbol{\lambda_3 F_1} + e_{32}$
- $y_{42} = \mu_{42} + \boldsymbol{\lambda_4 F_2} + e_{42}$
- $y_{52} = \mu_{52} + \boldsymbol{\lambda_5 F_2} + e_{52}$
- $y_{62} = \mu_{62} + \boldsymbol{\lambda_6 F_2} + e_{62}$
- **Both factor variances estimated** and a factor covariance, but both factor means are STILL fixed to 0

2. **Metric** Invariance Model

- Compare metric invariance to configural invariance model: **Is the model fit not worse** (−2ΔLL not significant)?
  - Check that factor variances are fixed to 1 in reference group only: they should be freely estimated in the other group, otherwise you are imposing a structural constraint (that groups have same variability) too
  - Otherwise, inspect the **modification indices** (voo-doo) to see if there are any indicators whose loadings want to differ across groups
  - Retest the model as needed after releasing one loading at a time, starting with the largest modification index, and continue until your partial metric invariance model is **not worse** than the configural model

- Do you have partial metric invariance (1+ loading per factor)?
  - Your trait is (sort of) measured in the same way across groups
  - If not, it doesn't make sense to evaluate how relationships involving the factor differ across groups (because the factor itself differs)
  - Even if full invariance holds, pry check the modification indices anyway

Step 3: Test "**Scalar**" Invariance

- Also called "**strong** factorial invariance"
- Do the groups have the same **indicator intercepts**?
  - Each indicator is still allowed to have a different intercept, but intercepts for same indicator are constrained equal across groups
  - Indicators that failed metric invariance do not get tested for scalar
  - Scalar invariance is required for factor mean comparisons!

- Previous (partial) metric invariance model is starting point, but change the method of model identification with respect to the intercepts and factor means: Estimate all newly constrained intercepts, but **fix the factor means to 0 in reference group** (free factor means in other group)
  - Why? Intercepts for marker items (if fixed=0 for identification) would be assumed invariant, and thus they could not be tested

3. **Scalar** Invariance Model:
Same Factor Loadings + Same Intercepts (saves +4 df)

Total DF across groups = 54 – 30 = 24 =

$$54 - \left(\mathbf{6\mu} + 12\sigma^2_{\mathrm{e}} + 6\lambda + 2\sigma^2_{\mathrm{F}} + 2\sigma_{\mathrm{F12}} + \mathbf{2\kappa_F}\right) = 24$$

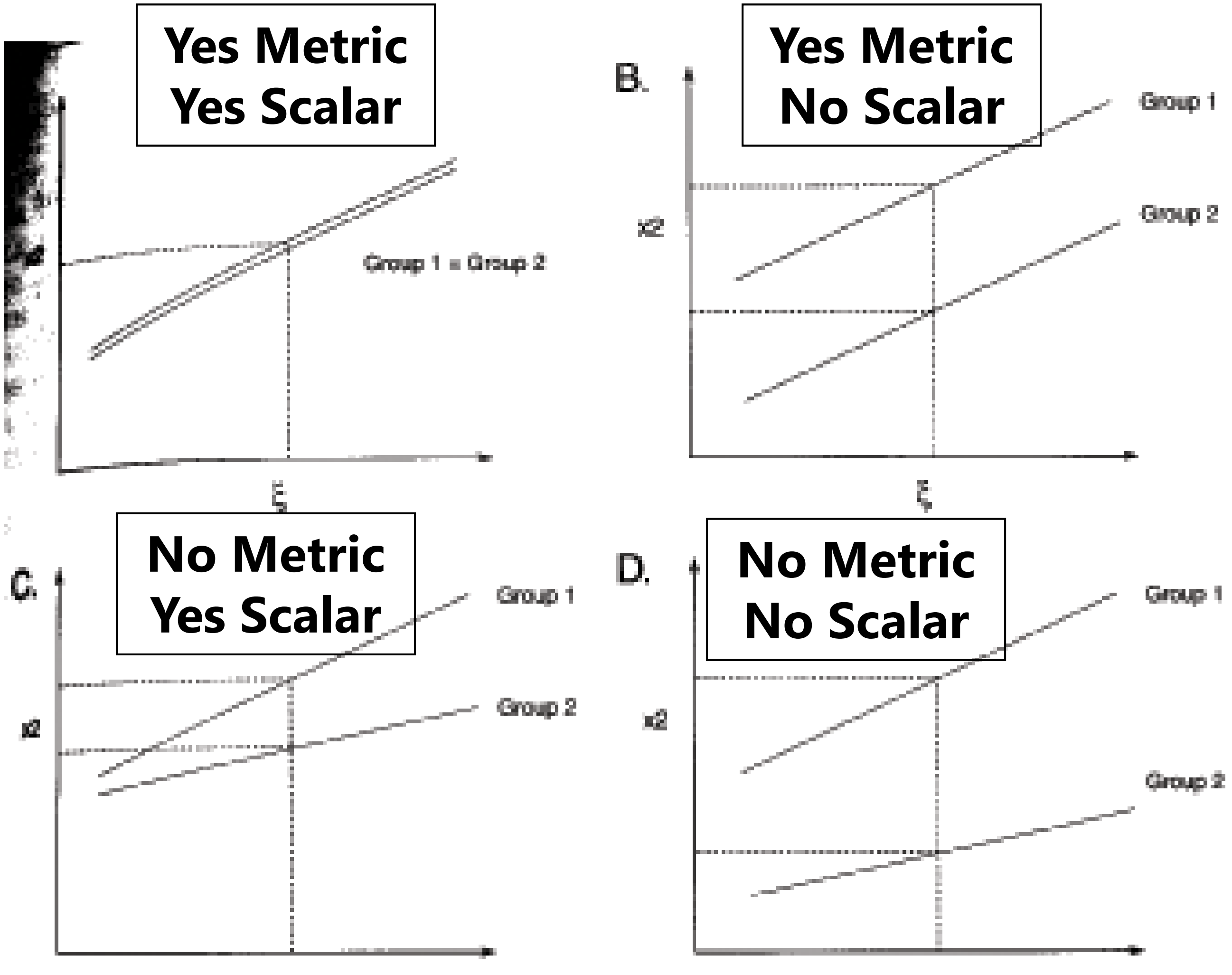
Group 1 (subscript = item, group):

- $y_{11} = \boldsymbol{\mu_1} + \lambda_1 F_1 + e_{11}$
- $y_{21} = \boldsymbol{\mu_2} + \lambda_2 F_1 + e_{21}$
- $y_{31} = \boldsymbol{\mu_3} + \lambda_3 F_1 + e_{31}$
- $y_{41} = \boldsymbol{\mu_4} + \lambda_4 F_2 + e_{41}$
- $y_{51} = \boldsymbol{\mu_5} + \lambda_5 F_2 + e_{51}$
- $y_{61} = \boldsymbol{\mu_6} + \lambda_6 F_2 + e_{61}$
- **Both factor variances fixed to 1, both factor means fixed to 0 for identification**, factor covariance is still estimated

Group 2 (subscript = item, group):

- $y_{12} = \boldsymbol{\mu_1} + \lambda_1 F_1 + e_{12}$
- $y_{22} = \boldsymbol{\mu_2} + \lambda_2 F_1 + e_{22}$
- $y_{32} = \boldsymbol{\mu_3} + \lambda_3 F_1 + e_{32}$
- $y_{42} = \boldsymbol{\mu_4} + \lambda_4 F_2 + e_{42}$
- $y_{52} = \boldsymbol{\mu_5} + \lambda_5 F_2 + e_{52}$
- $y_{62} = \boldsymbol{\mu_6} + \lambda_6 F_2 + e_{62}$
- **Both factor variances estimated, both factor means estimated to become mean differences,** and factor covariance is still estimated

Implications of Non-Invariance



Yes Metric
Yes Scalar

Yes Metric
No Scalar

No Metric
Yes Scalar

No Metric
No Scalar

Latent Factor

Latent Factor

Without metric invariance: Because unequal loadings implies non-parallel slopes, the intercept will differ as a result. The size of the difference depends on where theta=0.

This is why scalar invariance is often not tested if metric invariance fails for a given indicator.

3. **Scalar** Invariance Model

- Compare scalar invariance to last metric invariance model: **Is the model fit not worse** (−2ΔLL not significant)?
  - Check that factor means are fixed to 0 in reference group only: they should be freely estimated in the other group, otherwise you are imposing a structural constraint (groups have same means) too
  - Otherwise, inspect the **modification indices** (voo-doo) to see if there are any indicators whose intercepts want to differ across groups
  - Retest the model as needed after releasing one intercept at a time, starting with the largest modification index, and continue until your partial scalar invariance model is **not worse** than last metric model

- Do you have partial scalar invariance (1+ intercept per factor)?
  - Your trait is (sort of) responsible for mean differences across groups
  - If not, it doesn't make sense to evaluate factor means differs across groups (because something else is causing those differences)
  - Even if full invariance holds, pry check the modification indices anyway

Step 4: Test **Residual Variance** Invariance

- Also called "**strict** factorial invariance"
- Do the groups have the same **residual variances**?
  - Each indicator is still allowed to have a different residual variance (i.e., this is not a parallel items model), but residual variances for same indicator are constrained equal across groups
  - Indicators that failed scalar invariance do not get tested for residual variance invariance (by convention, although you could)
  - Residual invariance is of debatable importance, because it means that whatever is causing "not the factor" differs across groups
  - Equal residual variances are commonly misinterpreted to mean "equal reliabilities"—this is ONLY the case if the factor variances are the same across groups, too (stay tuned)
- This is the last step of "measurement invariance"

4. **Residual** Invariance Model:

+ Same Residual Variances (saves +6 df)

Total DF across groups = 54 – 24 = 30 =

$$54-\left(6\mu+\boldsymbol{6\sigma_{e}^{2}}+6\lambda+2\sigma_{F}^{2}+2\sigma_{F12}+2\kappa_{F}\right)=30$$

Group 1 (subscript = item, group):

- $y_{11} = \mu_1 + \lambda_1 F_1 + \mathbf{e_1}$
- $y_{21} = \mu_2 + \lambda_2 F_1 + \mathbf{e_2}$
- $y_{31} = \mu_3 + \lambda_3 F_1 + \mathbf{e_3}$
- $y_{41} = \mu_4 + \lambda_4 F_2 + \mathbf{e_4}$
- $y_{51} = \mu_5 + \lambda_5 F_2 + \mathbf{e_5}$
- $y_{61} = \mu_6 + \lambda_6 F_2 + \mathbf{e_6}$
- **Both factor variances fixed to 1, both factor means fixed to 0 for identification**, factor covariance is still estimated

Group 2 (subscript = item, group):

- $y_{12} = \mu_1 + \lambda_1 F_1 + \mathbf{e_1}$
- $y_{22} = \mu_2 + \lambda_2 F_1 + \mathbf{e_2}$
- $y_{32} = \mu_3 + \lambda_3 F_1 + \mathbf{e_3}$
- $y_{42} = \mu_4 + \lambda_4 F_2 + \mathbf{e_4}$
- $y_{52} = \mu_5 + \lambda_5 F_2 + \mathbf{e_5}$
- $y_{62} = \mu_6 + \lambda_6 F_2 + \mathbf{e_6}$
- **Both factor variances estimated, both factor means estimated to become mean differences,** and factor covariance is still estimated

4. **Residual Variance** Invariance Model

- Compare residual invariance to last scalar invariance model: **Is the model fit not worse** (−2ΔLL not significant)?
  - Otherwise, inspect the **modification indices** (voo-doo) to see if there are any indicators whose residual variances want to differ across groups
  - Retest the model after releasing one residual variance at a time, starting with the largest modification index, and continue until your partial residual invariance model is **not worse** than last scalar model
- Do you have partial residual variance invariance (1+ residual variance per factor)?
  - Your groups have the same amount of "not the factor" in each item (???)
  - Even if full invariance holds, pry check the modification indices anyway
  - Also assess any residual covariances across groups if you have those
- Your (partial) residual invariance model is the new baseline for assessing structural invariance…

Testing **Structural** Invariance

- Are the **factor variances** the same across groups? (+1 df/factor)
  - Fix the factor variance in the alternative group to 1 (as in the ref group)
  - Is model fit worse? If so, the groups differ in their factor variances

- Is the **factor covariance** the same across groups? (+1 df per pair)
  - Fix the factor covariances equal across groups, is model fit worse?
  - Factor correlation will only be the same across groups if the factor variances are the same, too (if factor variances differ, then factor covariance will, too)

- Are the **factor means** the same across groups? (+1 df/factor)
  - Fix the factor mean in the alternative group to 0 (as in the ref group)
  - Is model fit worse? If so, the groups differ in their factor means

- It is **not problematic** if structural invariance doesn’t hold
  - Given measurement invariance, this is a substantive issue about differences in the latent trait amounts and relations (and that’s ok)

Summary: Invariance Testing in CFA

- In CFA: Testing invariance has two distinct parts:
  - Measurement invariance: Is your construct being measured in the same way by the indicators across groups/time?
    - Hope for at least "partial" invariance… otherwise, game over
  - Structural invariance: Do your groups/times differ in their distribution and/or means of the construct? Let's find out!
    - Structural differences are real and interpretable differences given measurement invariance of the constructs

- In IFA: Still called testing invariance
  - Conducted similarly (but not exactly the same) in Mplus
- In IRT: Now called testing "differential item functioning"
  - With different names and rules, not directly tested in Mplus

Differential Item Functioning (DIF)

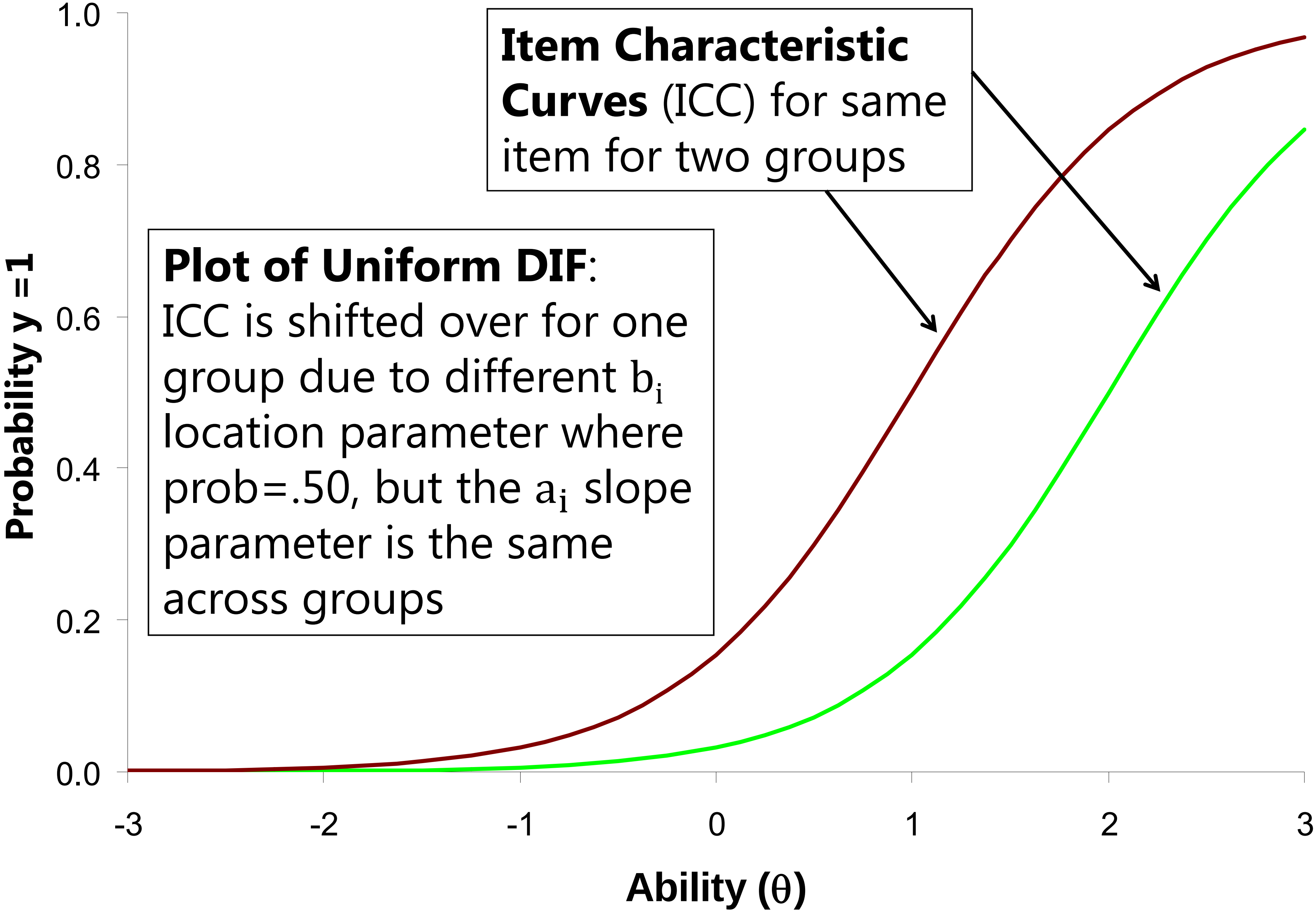
- In IRT (model with a_i discrimination and b_i difficulty), measurement NON-invariance = DIF
  - Note the inversion: Measurement Invariance = Non-DIF
    Measurement Non-Invariance = DIF
  - An item has "DIF" when persons with equal amounts of the traits, but from different groups, have different expected item responses
  - An item has "non-DIF" if persons with the same amount of the trait have the same expected item response, regardless of group
  - DIF can be examined across groups, over time, over conditions, etc., the same as in CFA/IFA
  - Independent groups? Multiple-group model
  - Dependent "groups"? One factor per "group" in same model

2 Types of DIF (as described in IRT)

- "**Uniform DIF**" → Analogous to scalar NON-invariance
  - IRT b_i parameters differ across groups
  - Item is systematically more difficult/severe for members of one group, *even for persons with the same amount of the theta trait*
  - Example: "I cry a lot" → Would men and women *with the same amount of depression* have the same expected item response?

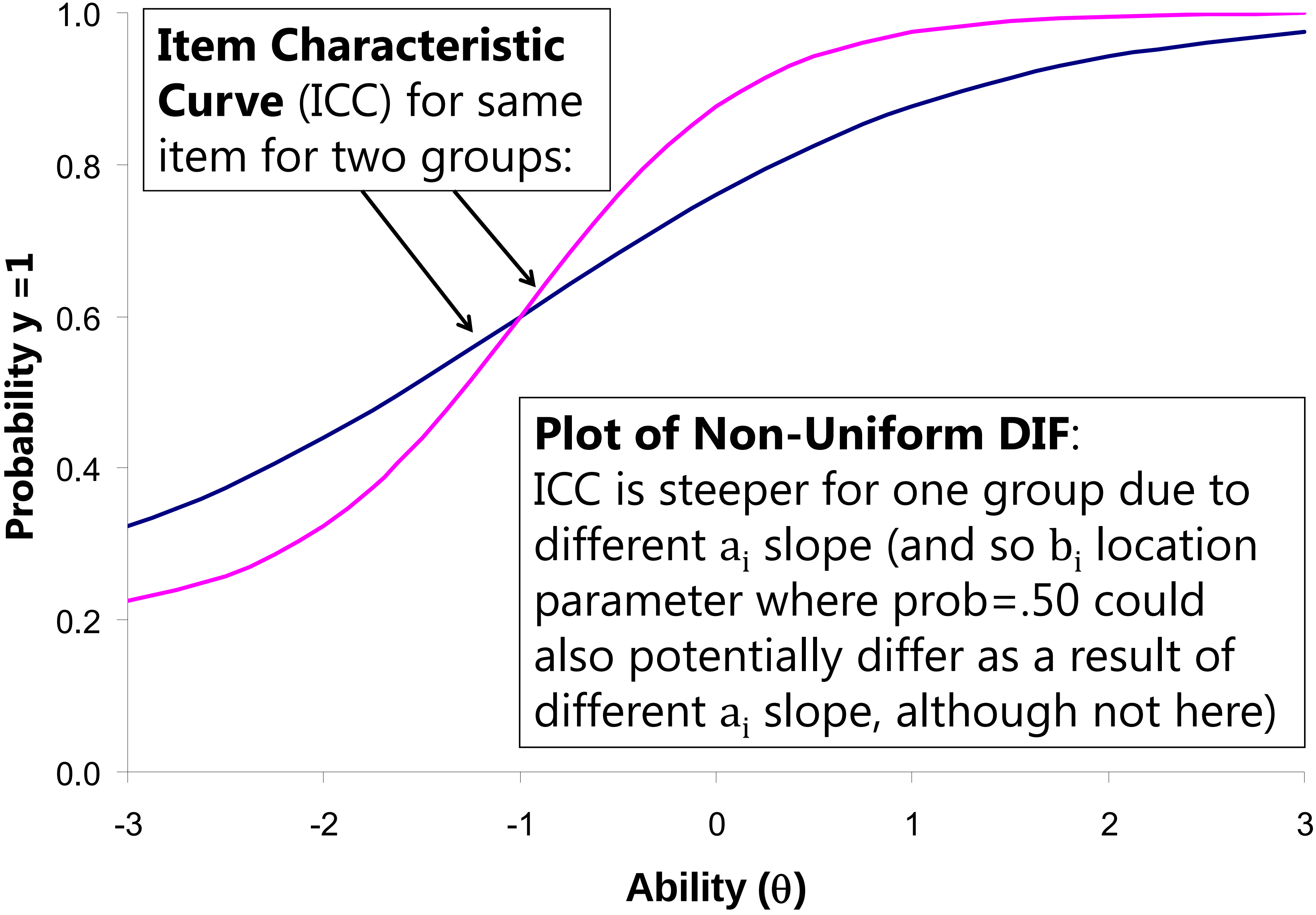
- "**Non-Uniform DIF**" → Analogous to metric NON-invariance
  - IRT a_i (and possibly b_i) parameters differ across groups
  - Item is systematically more related to theta for members of one group → higher discrimination (item "works better")
  - Shift in item difficulty is not consistent across theta continuum

- What about residual variance invariance? It depends:
  - Doesn't exist in ML: no estimated error variance (is logit=3.29 or probit=1.00)
  - Will exist in WLSMV after constraining loadings and thresholds, but not before…



Item Characteristic Curves (ICC) for same item for two groups

Plot of Uniform DIF: ICC is shifted over for one group due to different b_i location parameter where prob=.50, but the a_i slope parameter is the same across groups



Item Characteristic Curve (ICC) for same item for two groups:

Plot of Non-Uniform DIF:
ICC is steeper for one group due to different a_i slope (and so b_i location parameter where prob=.50 could also potentially differ as a result of different a_i slope, although not here)

Testing Measurement Invariance in Categorical Outcomes

- 2 versions of model for polytomous outcomes in Mplus:
  - IRT: $\text{Logit or Probit}(y_{kis} = 1) = a_i(\theta_s - b_{ki})$
  - IFA: $\text{Logit or Probit}(y_{kis} = 1) = -\tau_{ki} + \lambda_i\theta_s$
    - Logit or Probit in ML; only Probit in WLSMV

The k thresholds divide the C item responses into $C - 1$ cumulative binary submodels ($y = 0$ if lower, $y = 1$ if higher)

- Mplus estimates the IFA τ_{ki} and λ_i parameters, then *converts* to the IRT a_i and b_{ki} parameters for binary (but not polytomous) outcomes
  - Tests of measurement invariance are specifically for τ_{ki} and λ_i, not a_i and b_{ki}
  - So Mplus does not directly allow examination of "DIF" for a_i and b_i directly
- **IFA τ_{ki} and λ_i are held directly invariant, not IRT a_i and b_i**
  - So even if λ_i factor loadings are invariant across groups, IRT a_i discriminations will still differ across groups due to differences in their theta variances
  - Likewise, even if τ_{ki} thresholds are invariant across groups/time, IRT b_i difficulty parameters can still differ due to differences in theta mean and theta variance

Review: From IFA to IRT

IFA with "easiness" **intercept $\boldsymbol{\mu_i}$**: $\textbf{Logit or Probit } \mathbf{y_{is}} = \boldsymbol{\mu_i} + \boldsymbol{\lambda_i}\mathbf{F_s}$ $\quad \boldsymbol{\mu_i} = -\boldsymbol{\tau_i}$

IFA with "difficulty" **threshold $\boldsymbol{\tau_i}$**: $\textbf{Logit or Probit } \mathbf{y_{is}} = -\boldsymbol{\tau_i} + \boldsymbol{\lambda_i}\mathbf{F_s}$

IFA model with "difficulty" thresholds can be written as a **2-PL IRT Model:**

IRT model: **IFA model:**

$$\textbf{Logit or Probit } \mathbf{y_{is}} = \mathbf{a_i}(\boldsymbol{\theta_s} - \mathbf{b_i}) = \underbrace{-\mathbf{a_i}\mathbf{b_i}}_{\boldsymbol{\tau_i}} + \underbrace{\mathbf{a_i}}_{\boldsymbol{\lambda_i}}\boldsymbol{\theta_s}$$

$\mathbf{a_i}$ = discrimination
$\mathbf{b_i}$ = difficulty
$\boldsymbol{\theta_s}$ = F_s latent trait

Convert IFA to IRT:

$$a_i = \lambda_i * \sqrt{\text{Theta Variance}}$$

$$b_i = \frac{\tau_i - (\lambda_i * \text{Theta Mean})}{\lambda_i * \sqrt{\text{Theta Variance}}}$$

Convert IRT to IFA:

$$\lambda_i = \frac{a_i}{\sqrt{\text{Theta Variance}}}$$

$$\tau_i = a_i b_i \frac{\text{Theta Mean}}{\sqrt{\text{Theta Variance}}}$$

Note: prior to Mplus v7, these formulas will differ when using logit or probit

Invariance Testing in Mplus

- **IFA using Full-Information MML:** Multiple group models are not permitted, but you can trick Mplus into doing it (e.g., here, by group):
  - VARIABLE: KNOWNCLASS = group (men=1, women=2);
  - ANALYSIS: TYPE = MIXTURE;
  - MODEL: %OVERALL% (… model for reference group listed here)
    %group#2% (… model for alternative group goes here)

- **IFA using Limited-Information WLSMV:** Mplus does allow multiple group models, with a few useful other benefits
  - Faster estimation if you have multiple factors/thetas
  - DIFFTEST does nested model comparisons for you (still going for “not worse”)
  - Can get modification indices (voo-doo) to troubleshoot non-invariance
  - Can test differences in residual variances (in THETA parameterization)

- In either method, the same category responses must be observed for each group, otherwise you cannot test the item thresholds

Configural Invariance Baseline Model for Categorical Outcomes (2 Groups)

- **Factor variances**: fixed to 1 in both groups
- **Factor covariances**: if any, free in both groups
- **Factor means**: fixed to 0 in both groups

You could also use the same configural model identification as in CFA (your choice of scale)

- **Factor loadings**: all freely estimated (so each can be tested later)
  - Remember: IRT a_i parameters will still vary across groups even after loadings are constrained because of group differences in theta variance
- **Item Thresholds**: all freely estimated (so each can be tested later)
  - Remember: IRT b_{ki} parameters will still vary across groups even after thresholds are constrained because of group differences in theta mean and theta variance
- **Fix all residual variances=1 <u>in all groups</u>**
  - Groups will eventually be allowed to differ in both factor variance and “error variance” (proxy for total variation in WLSMV models)

Sequential Invariance Models

Note: Save for DIFFTEST at each step!

- **Step 1**: Fit baseline configural invariance model across groups
  - Should be "close enough" factor structures, otherwise game over
  - Alternative group is allowed different loadings and thresholds, SAME residual variances=1
- **Step 2 (Metric-ish)**: Constrain all loadings equal but free factor variances in alternative group—is fit worse relative to configural model?
  - If fit is worse, check MODINDICES to see why; release problematic constrained loadings <u>one at a time</u>; check fit against configural model to see if it's not worse yet
- **Step 3 (Scalar-ish)**: Constrain thresholds equal for items that passed metric but free factor means in alternative group—is fit worse relative to metric model?
  - If fit is worse, check MODINDICES to see why; release problematic constrained thresholds <u>one item at a time</u>; check fit against metric model to see if it's not worse yet
  - MODINDICES may want the "intercept" free, but this is not possible to do, so focus on problematic (non-invariant) item thresholds instead
  - Reasonable people disagree: Mplus recommends doing steps 2 and 3 in one step because loadings and thresholds are dependent; others disagree (see Millsap's 2011 book; all of IRT)

Sequential Invariance Models

Note: Save for DIFFTEST at each step!

- **Step 4**: Test if residual variances for items that passed scalar in alternative group ≠1 → differ from reference group (in which residual variance = 1)
  - Differences in residual variances between groups are not identified until you have at least some of the loadings and thresholds constrained across groups
  - Consequently, this test proceeds backwards: first estimated is the “bigger” non-invariant residual variance model, second estimated is the “smaller” original scalar invariance model (in which residual variances were fixed to 1 for all items for all groups)
  - Differential residual variances can be a proxy for group differences in overall variability, but this model may not always converge (if it doesn’t, skip this step, but note doing so)

- **Steps 5, 6, 7:** Test Structural Invariance (just like before in CFA):
  - Constrain equal across groups in sequential models: factor variances, then factor covariances, and then factor means (equal to 0) to test for “real” group differences
  - Same story as in CFA: Only if you have at least partial measurement invariance can structural group/time/condition differences be meaningfully interpreted

- Factors are the same no matter what measurement model was used to create them… so now we are ready to use them to do SEM!