Methods for Intersectional Measurement Invariance Testing

Dakota W. Cintron
Cornell University

Modern Modeling Methods Conference
June 27, 2023

Aims

Provide an overview on using three methods for performing *intersectional* measurement invariance testing:

1) alignment method
2) mixture multiple group factor analysis
3) moderated nonlinear factor analysis

Demonstrate methods with an empirical example

Measurement Invariance

Measurement invariance assesses the psychometric equivalence of a construct across groups or across time

When psychological tests are used in diverse populations, it is assumed that a given test score represents the same level of the underlying construct (e.g., achievement or depression) across groups

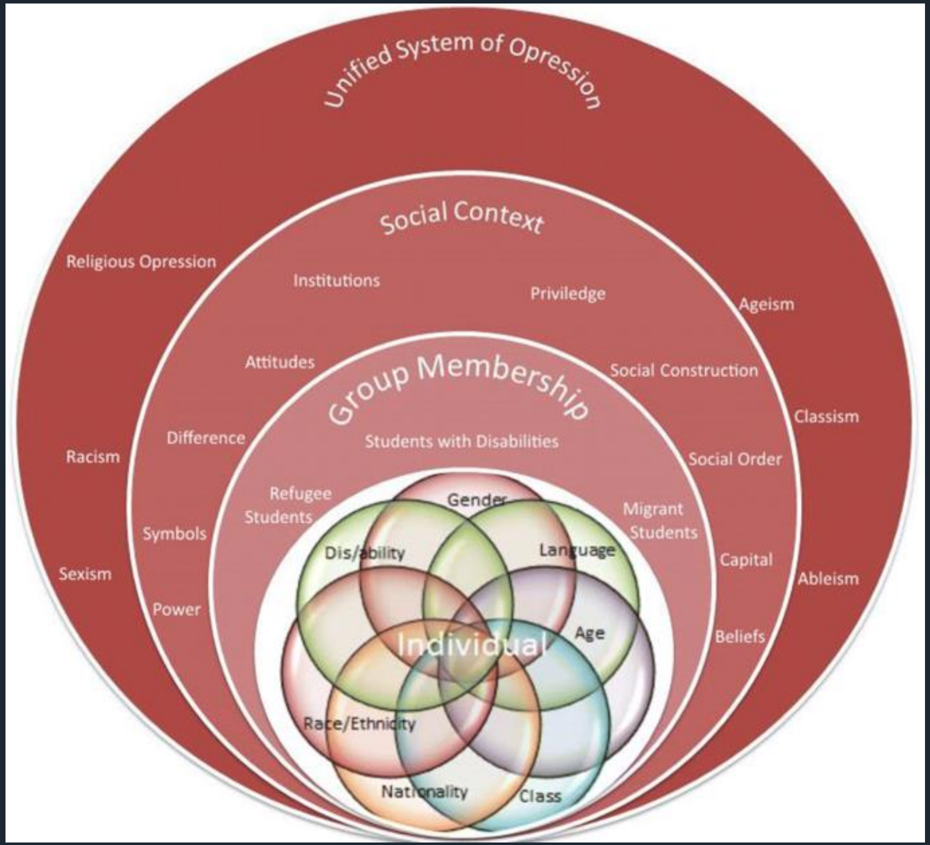
Measurement noninvariance suggests that a construct has a different structure or meaning to different groups or on different measurement occasions in the same group, and so the construct cannot be meaningfully tested or construed across groups or across tim

Intersectionality Theory

Kimberlé Crenshaw (1989) coined the idea when arguing for an intersectional legal framework for examining how race and gender interact to shape Black women's employment experiences and exposure to discrimination

In psychological and epidemiological research, intersectionality theory is used as a framework for understanding how multiple social or political identities (e.g., gender, age, or race) intersect to affect an individual's lived experience and health outcomes as opposed to considering each factor disparately

May require many potential intersectional subgroups



Bešić, 2020

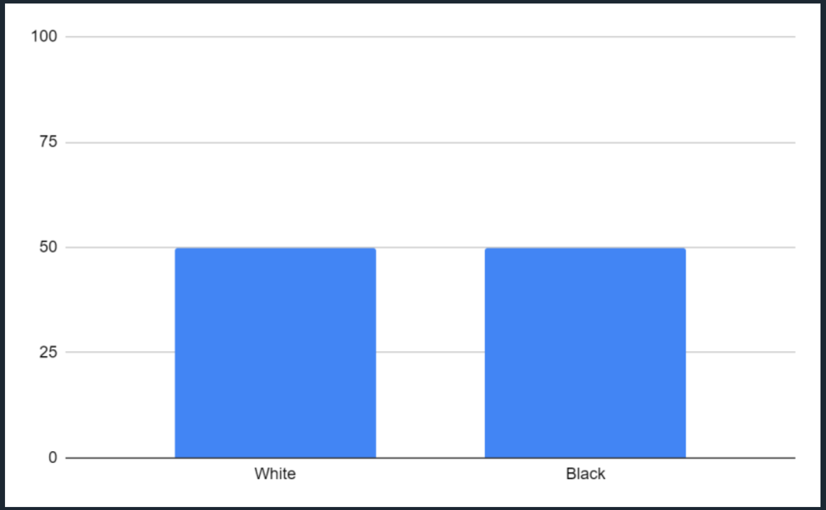
Problem

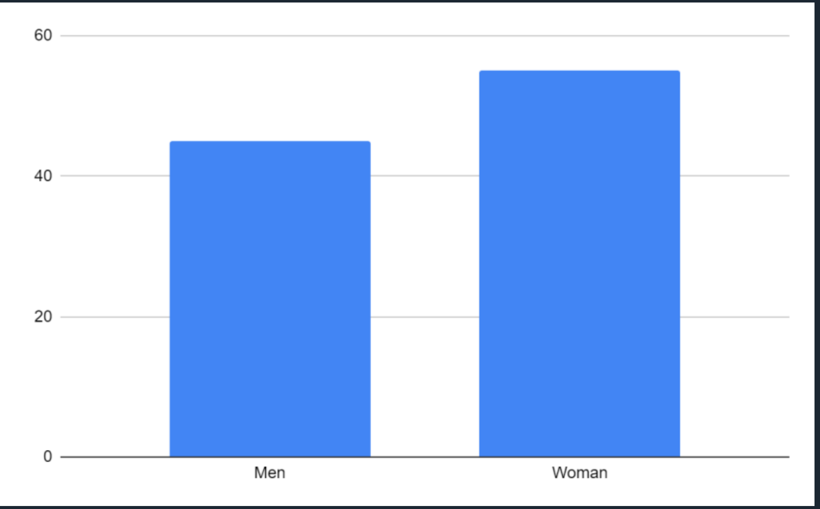
Assessing measurement invariance is crucial to help determine whether group differences in a latent construct (e.g., depression) are meaningful or not

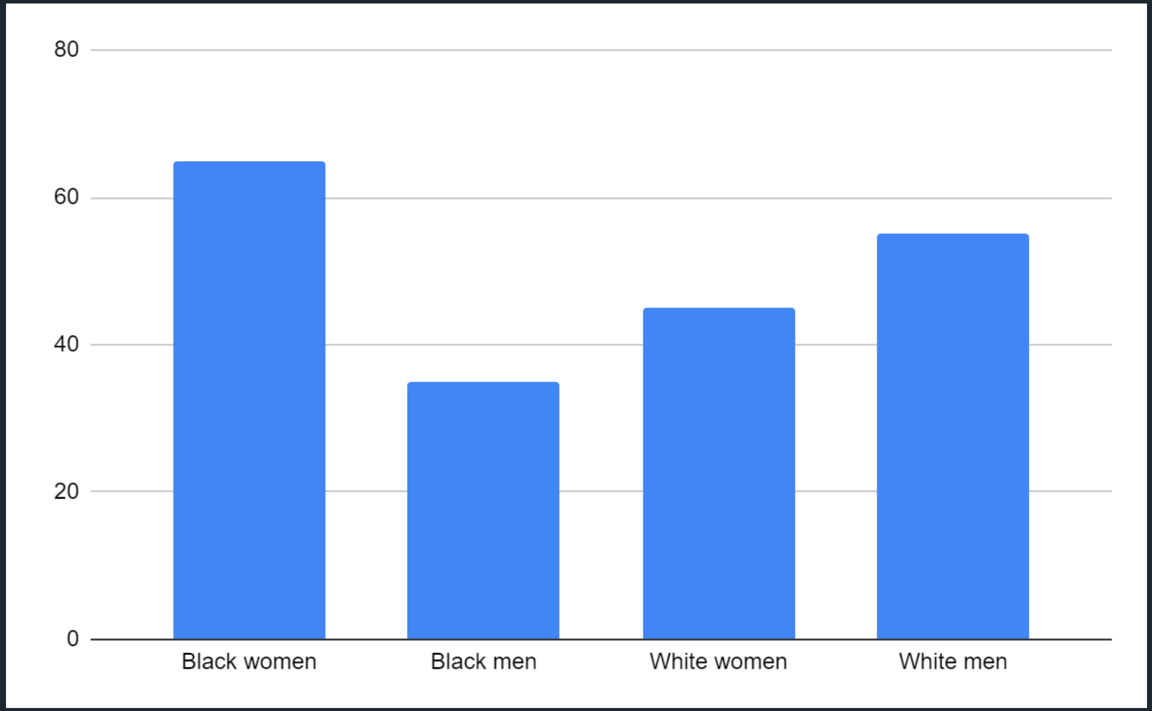
Traditionally, the evaluation of measurement invariance involves one demographic variable (e.g., gender) with a few subgroups (male, female)

***Intersectional* measurement invariance testing requires evaluating the psychometric properties of a scale across potentially many social and political identities**

Simple Example of Ignoring Intersectionality





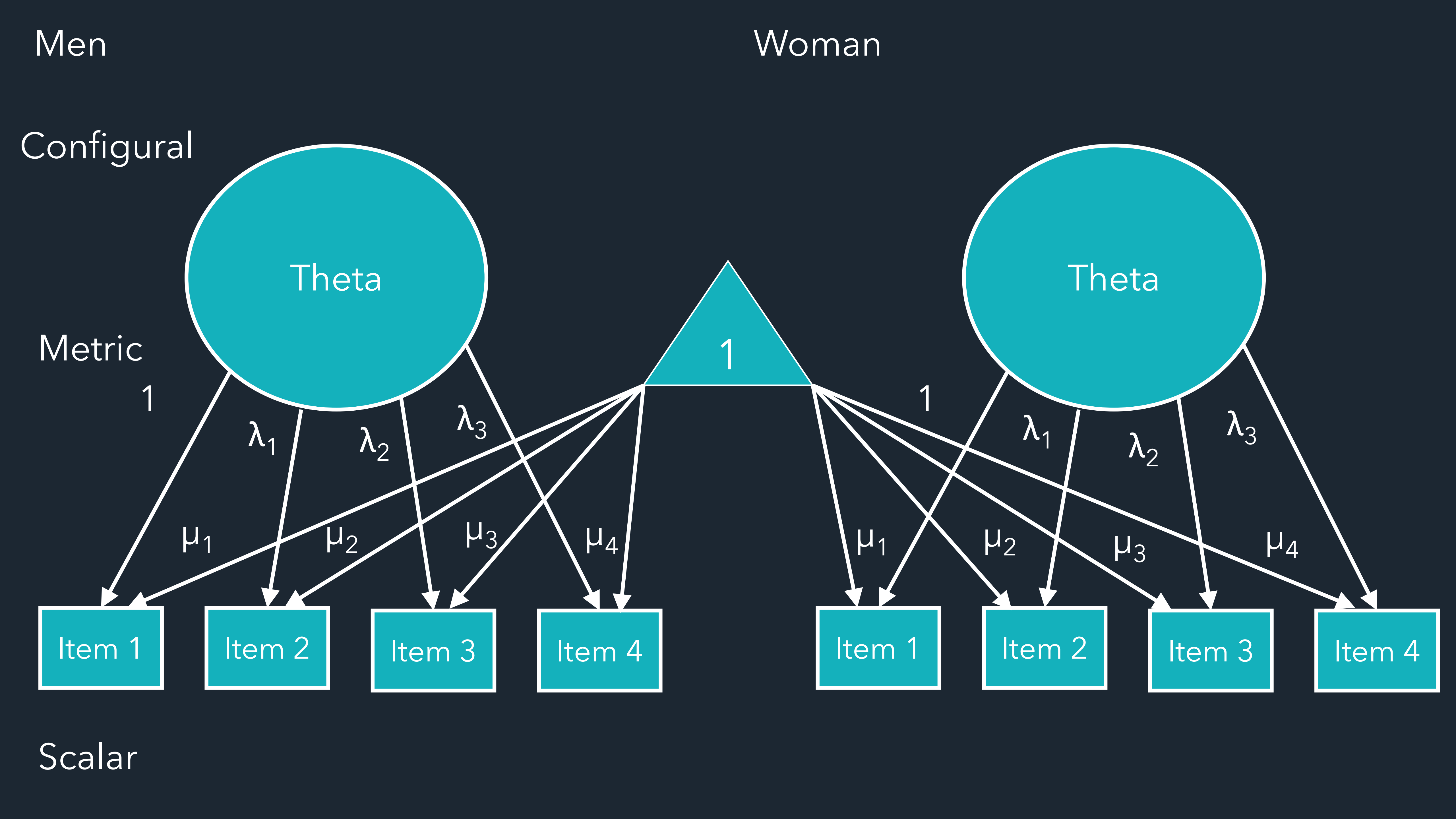


Traditional Measurement Invariance Testing

- Measurement invariance (MI) testing (Meredith, 1993) involves comparing configural, metric, and scalar models using multiple-group confirmatory factor analysis (CFA)
  - Configural invariance implies the same factor structure across groups
  - Metric invariance implies the factor loadings are invariant across groups
  - Scalar invariance implies that the factor loadings and thresholds (item intercepts) are invariant across groups
    - Scalar invariance is often considered a requirement for comparing factor means between groups
- Configural, metric, and invariance tests are generally conducted using a sequential constraint imposition or forward approach (see Horn & McArdle, 1992)
  - Compare the model fit of configural to metric, and then metric to scalar

Challenges with Traditional MI Testing

The evaluation of MI is generally limited to comparing independent groups defined by a single demographic variable


Men
Woman
Configural
Metric
Scalar
Theta
Theta
1
1
1
λ_1
λ_2
λ_3
λ_1
λ_2
λ_3
μ_1
μ_2
μ_3
μ_4
μ_1
μ_2
μ_3
μ_4
Item 1
Item 2
Item 3
Item 4
Item 1
Item 2
Item 3
Item 4

Intersectional Measurement Invariance

- It is important to consider intersectionality theory for measurement invariance testing because it is possible that the intersection of individuals identities may shape their response behavior
- Intersectional measurement invariance testing requires evaluating the psychometric properties of a scale across potentially many social and political identities (Cintron et al., 2023)
    - The intersection of race (White, Hispanic, Black, and Asian), gender (male, female), education (high school, bachelor's degree, graduate), and economic advantage (disadvantaged, advantaged) would result in 4x 2 x 3 x 2 = 48 intersectional subgroups.
- May require many potential subgroups and thus requires measurement invariance testing methods that can handle many groups
    - Alignment method
    - Mixture multiple-group factor analysis
    - Moderated nonlinear factor analysis

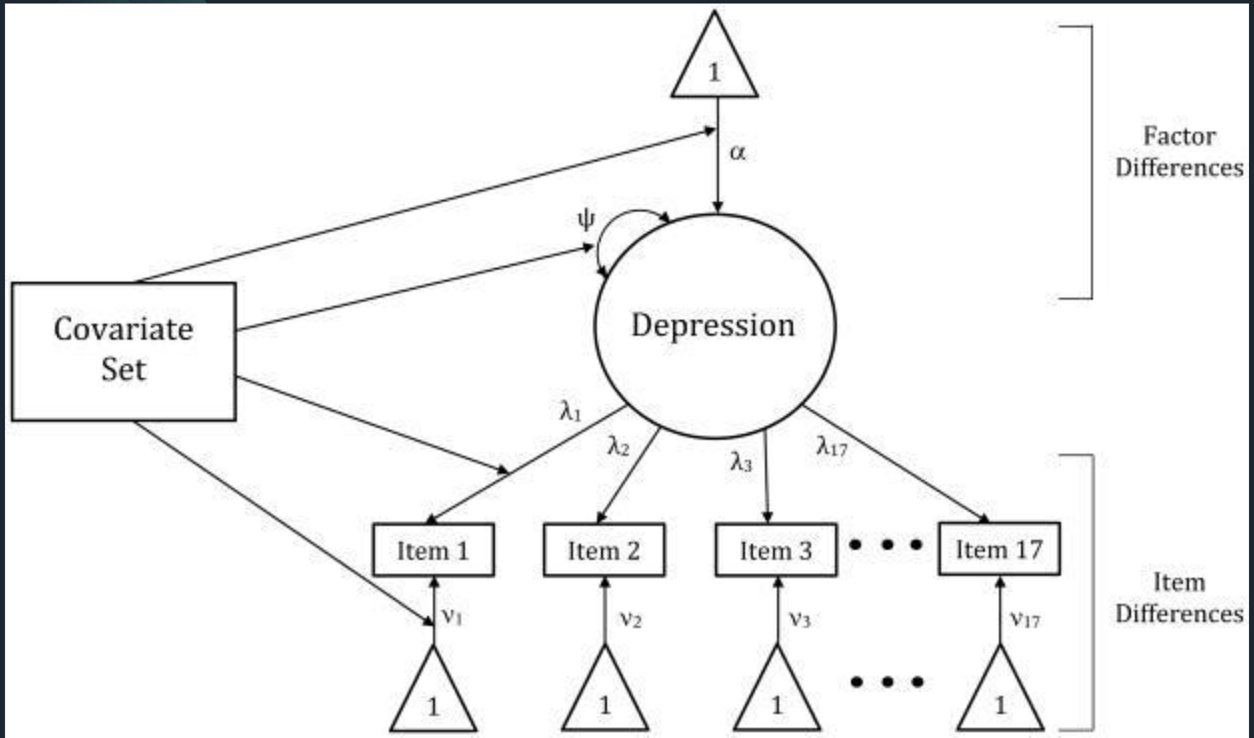
Alignment Method

- Asparouhov and Muthén (2014) developed the alignment method as an alternative to traditional multiple group CFA approaches for data structures with many groups
  - Can accommodate two or more groups easily
- The alignment method aims to make unbiased factor mean comparisons by producing a factor model with factor loadings and item intercepts that are as close to equivalent as possible across groups
  - In other words, the alignment approach aims to minimize measurement non-invariance
- Not necessarily a measurement invariance testing procedure but rather an *optimization approach* for finding the optimal amount of measurement invariance
  - See Luong and Flake (2021) and Asparouhov & Muthen (2014) for more details
  - Assumes configural invariance
- We implement the alignment method in *Mplus*, see also *sirt* package in R

Mixture Multiple Group Factor Analysis

- With many groups, it is unlikely that all groups have evidence of scalar invariance but equally unlikely that each group has its own unique set of parameters (loadings, intercepts)
- Mixture multigroup factor analysis (MMG-FA; De Roover, 2021) clusters groups according to a specific level of measurement invariance
    - Groups with scalar invariance are obtained by imposing cluster-specific intercepts and invariant loadings whereas unique variances, factor means, and factor (co)variances can differ between groups
    - For each cluster where scalar invariance holds, latent mean comparisons may be made
    - Uses three methods to choose the number of clusters BIC, scree rations, and convex hull of the likelihood
    - Assumes configural invariance
- We implement MMG-FA using the *mixmgfa* package in R

Moderated Nonlinear Factor Analysis (MNLFA)



Curran et al. (2014)

- Measurement invariance is assessed through parameter moderation in a single-group CFA model (Curran et al., 2014)
  - Can handle the assessment of measurement invariance across multiple continuous and categorical background variables
- A CFA model is estimated in which background variables are included as moderated variables
  - All parameters may be moderated by the background variables
- Nonzero effects of the covariates on the parameter's indicative of violations of invariance
  - Linear and nonlinear relationships possible
- Assumes configural invariance
- We implement MNLFA in *Mplus* using the R package *aMNLFA*

Simulation Example

Two Conditions

- Simulate data for 8 groups
- Invariant model

```
# model generating functions
invariant_model <- 'f1=~x1+.6*x2+.7*x3+.5*x4 + 0.9*x5 + 0.8*x6 + 0.6*x7
                    f2=~x8+.5*x9+.8*x10+.9*x11 + 0.7*x12 + 0.6*x13 + 0.7*x14
                    f1 ~~ 0.35*f2'
```

- Noninvariant model
  - Medium-large factor loading noninvariance for items 3, 7, and 13 for groups 6-8
  - Medium item intercept noninvariance for items 4, 5, and 14 for groups 6-8
    - Constant of 0.5 to items

```
noninvariant_model <- 'f1=~x1+.6*x2+.35*x3+.5*x4 + 0.9*x5 + 0.8*x6 + 0.25*x7
                       f2=~x8+.5*x9+.8*x10+.9*x11 + 0.7*x12 + 0.25*x13 + 0.7*x14
                       f1 ~~ 0.35*f2'
```

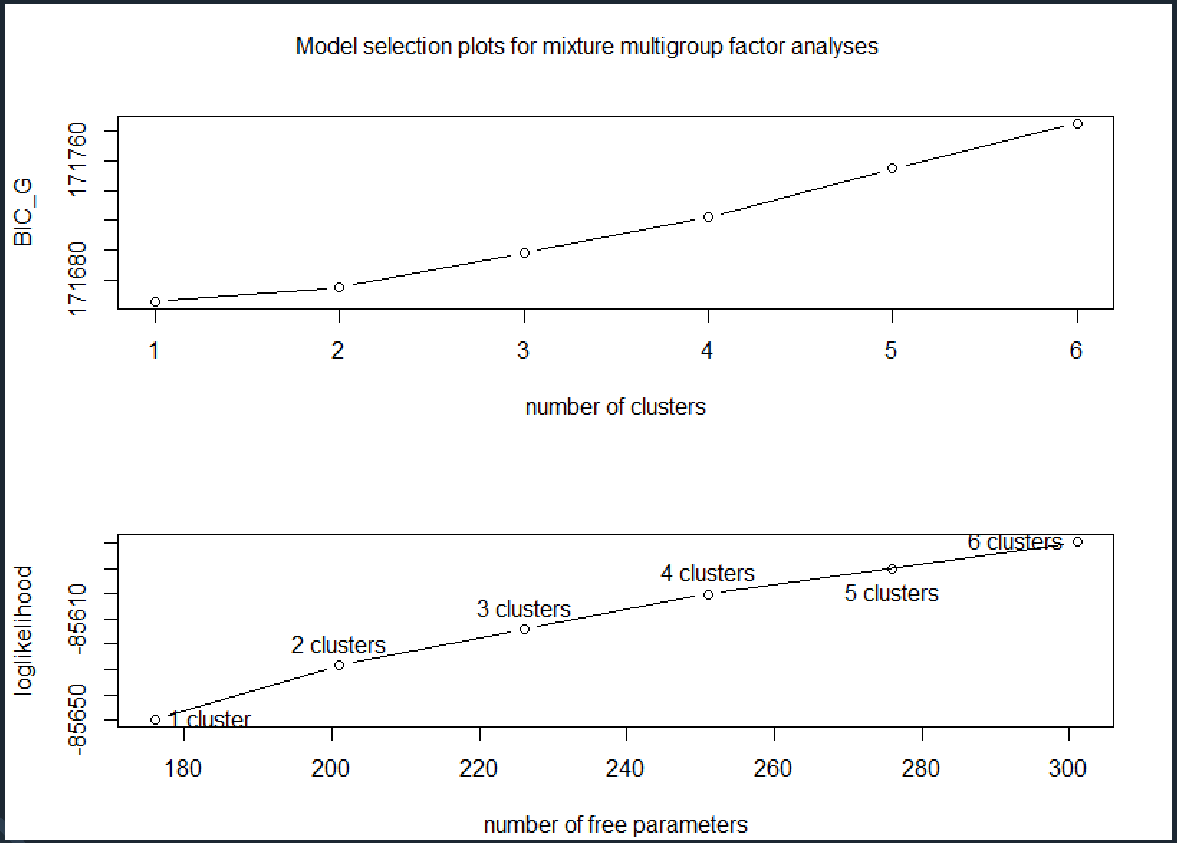
Results: Traditional MG-CFA (Invariant)

```
################### Nested Model Comparison #########################

Chi-Squared Difference Test

                Df    AIC    BIC  Chisq Chisq diff     RMSEA Df diff Pr(>Chisq)
fit_configural 608 171784 173949 602.65
fit_metric     692 171702 173338 688.39     85.733 0.0064239      84     0.4269
fit_scalar     776 171632 172740 786.38     97.988 0.0182498      84     0.1411

####################### Model Fit Indices ###########################
                  chisq  df pvalue rmsea    cfi    tli  srmr         aic
fit_configural 602.655† 608   .554 .000† 1.000† 1.001† .029† 171784.219
fit_metric     688.388  692   .532 .000† 1.000† 1.000  .036  171701.952
fit_scalar     786.376  776   .390 .005  0.999  0.999  .038  171631.940†
                    bic
fit_configural 173949.372
fit_metric     173338.405
fit_scalar     172739.693†

################## Differences in Fit Indices #######################
                          df rmsea    cfi    tli  srmr     aic      bic
fit_metric - fit_configural 84 0.000  0.000  0.000 0.007 -82.267 -610.967
fit_scalar - fit_metric     84 0.005 -0.001 -0.001 0.002 -70.012 -598.712
```

Results Alignment (Invariant)

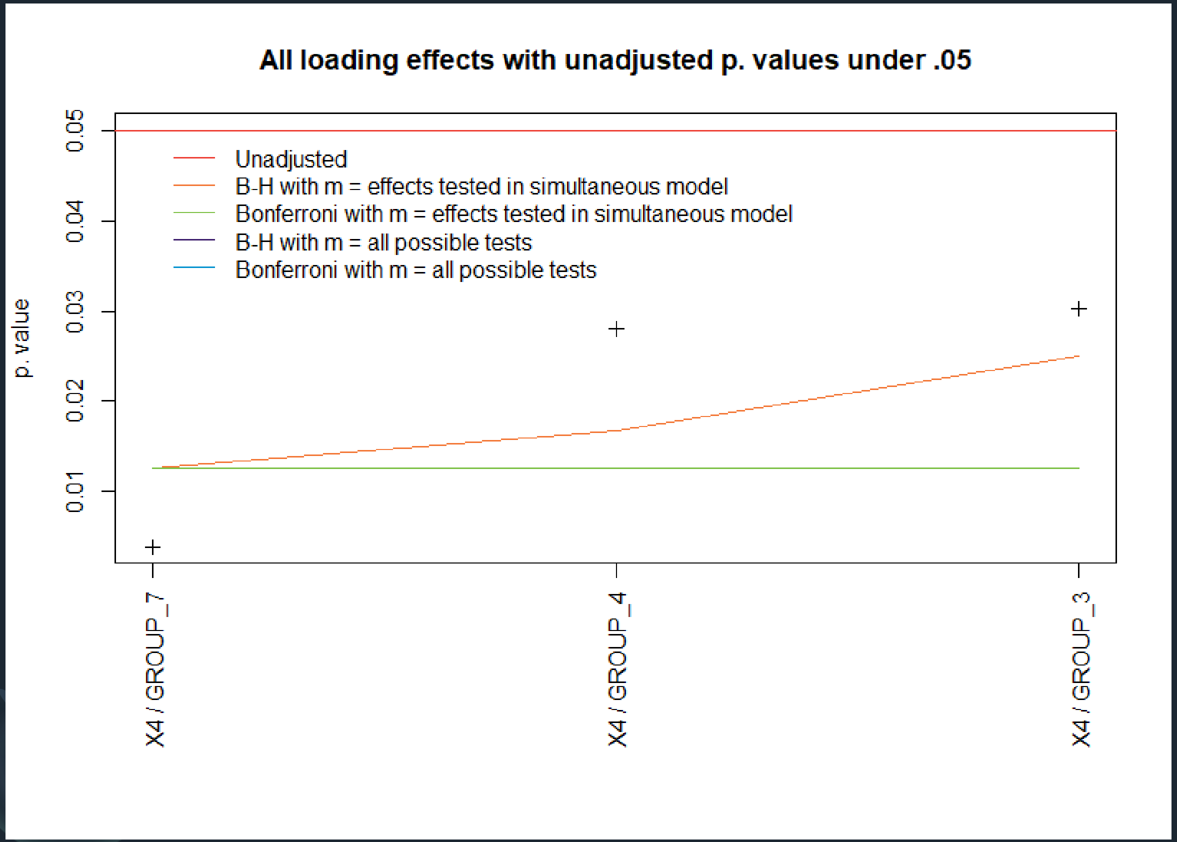
```
Intercepts/Thresholds
  X1           1 2 3 4 5 6 7 8
  X2           1 2 3 4 5 6 7 8
  X3           1 2 3 4 5 6 7 8
  X4           1 2 3 4 5 6 7 8
  X5           1 2 3 4 5 6 7 8
  X6           1 2 3 4 5 6 7 8
  X7           1 2 3 4 5 6 7 8
  X8           1 2 3 4 5 6 7 8
  X9           1 2 3 4 5 6 7 8
  X10          1 2 3 4 5 6 7 8
  X11          1 2 3 4 5 6 7 8
  X12          1 2 3 4 5 6 7 8
  X13          1 2 3 4 5 6 7 8
  X14          1 2 3 4 5 6 7 8

Loadings for F1
  X1           1 2 3 4 5 6 7 8
  X2           1 2 3 4 5 6 7 8
  X3           1 2 3 4 5 6 7 8
  X4           1 2 3 4 5 6 7 8
  X5           1 2 3 4 5 6 7 8
  X6           1 2 3 4 5 6 7 8
  X7           1 2 3 4 5 6 7 8

Loadings for F2
  X8           1 2 3 4 5 6 7 8
  X9           1 2 3 4 5 6 7 8
  X10          1 2 3 4 5 6 7 8
  X11          1 2 3 4 5 6 7 8
  X12          1 2 3 4 5 6 7 8
  X13          1 2 3 4 5 6 7 8
  X14          1 2 3 4 5 6 7 8
```

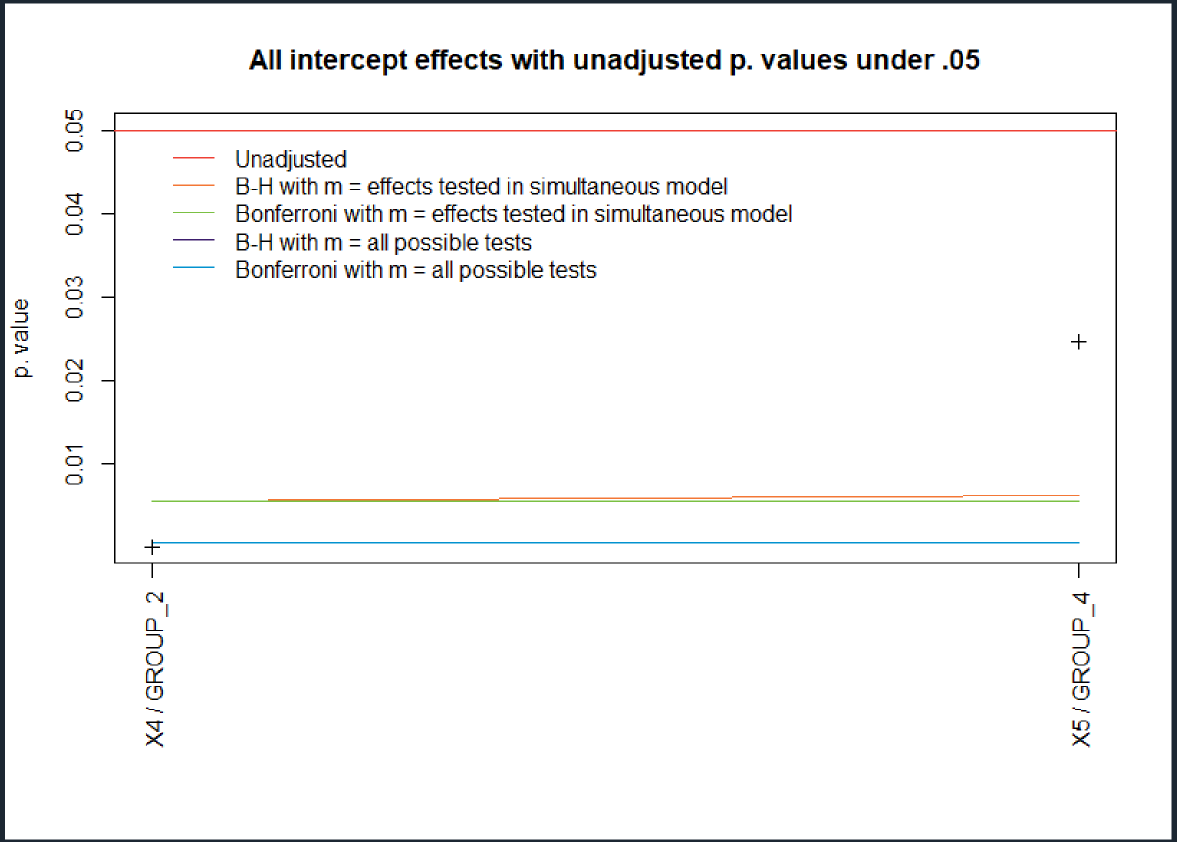
Results: MMG-FA (Invariant)



	Factor_1	Factor_2
x1	1.0230432	0.0000000
x2	0.6031932	0.0000000
x3	0.6996908	0.0000000
x4	0.5208666	0.0000000
x5	0.9350079	0.0000000
x6	0.8018651	0.0000000
x7	0.6276809	0.0000000
x8	0.0000000	0.9756335
x9	0.0000000	0.4732969
x10	0.0000000	0.7916525
x11	0.0000000	0.8731683
x12	0.0000000	0.7132992
x13	0.0000000	0.5636934
x14	0.0000000	0.7052766

Results: MNLFA (Invariant)





Results: Traditional MG-CFA (Noninvariant)

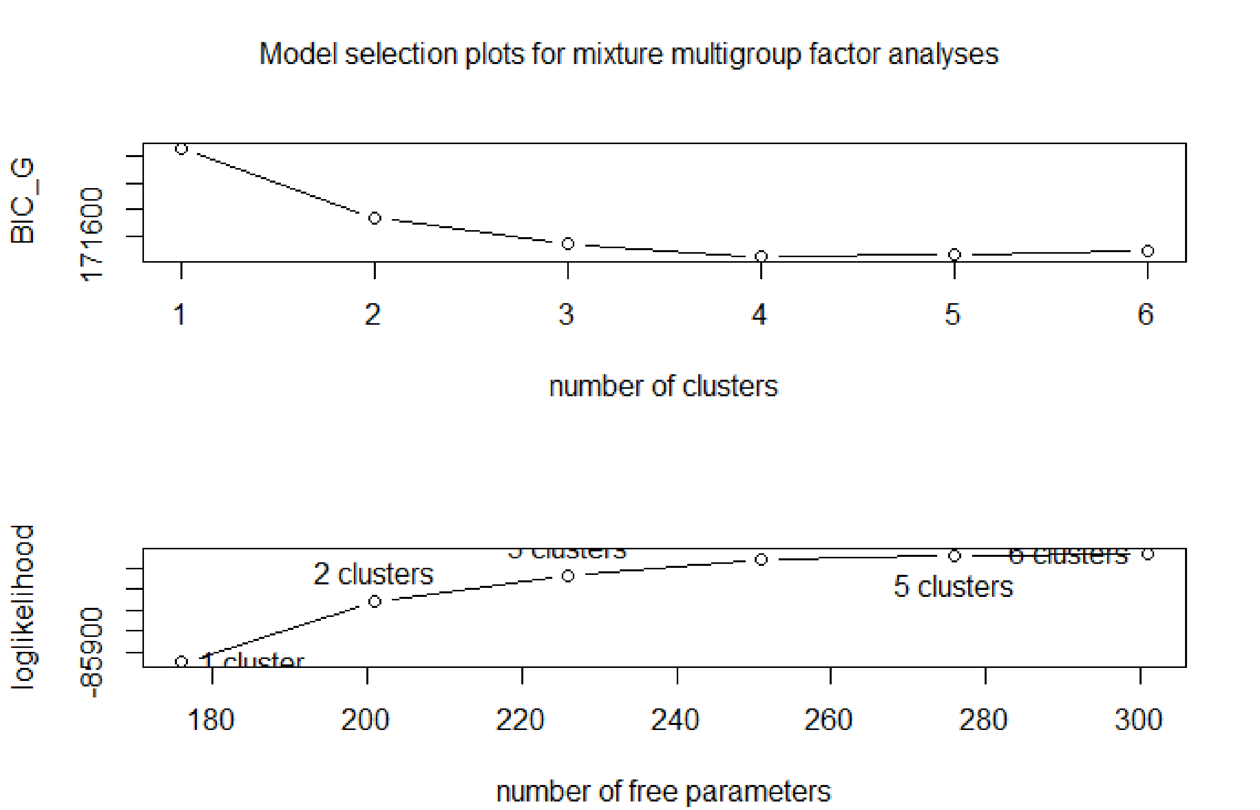
```
################### Nested Model Comparison #########################

Chi-Squared Difference Test

                 Df    AIC    BIC   Chisq Chisq diff    RMSEA Df diff Pr(>Chisq)
fit_configural 608 171479 173644  599.85
fit_metric     692 171629 173265  917.81     317.96 0.074635      84  < 2.2e-16
fit_scalar     776 172099 173206 1555.75     637.94 0.114844      84  < 2.2e-16

fit_configural
fit_metric     ***
fit_scalar     ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

####################### Model Fit Indices ###########################
                   chisq  df pvalue rmsea    cfi    tli  srmr         aic
fit_configural  599.850† 608   .585 .000† 1.000† 1.001† .029† 171478.832†
fit_metric      917.808  692   .000 .026   .980   .979  .050  171628.790
fit_scalar     1555.752  776   .000 .045   .930   .934  .059  172098.734
                       bic
fit_configural 173643.985
fit_metric     173265.243
fit_scalar     173206.486†

################## Differences in Fit Indices #######################
                            df rmsea    cfi    tli  srmr     aic      bic
fit_metric - fit_configural 84 0.026  -0.02 -0.022 0.021 149.958 -378.742
fit_scalar - fit_metric     84 0.019  -0.05 -0.044 0.009 469.944  -58.756
```

Results: Alignment (Noninvariant)

```
Intercepts/Thresholds
  X1           1 2 3 4 5 6 7 8
  X2           1 2 3 4 5 6 7 8
  X3           1 2 3 4 5 6 7 8
  X4           1 2 3 4 5 (6) (7) (8)
  X5           1 2 3 4 5 (6) (7) (8)
  X6           1 2 3 4 5 6 7 8
  X7           1 2 3 4 5 6 7 8
  X8           1 2 3 4 5 6 7 8
  X9           1 2 3 4 5 6 7 8
  X10          1 2 3 4 5 6 7 8
  X11          1 2 3 4 5 6 7 8
  X12          1 2 3 4 5 6 7 8
  X13          1 2 3 4 5 6 7 8
  X14          1 2 3 4 5 (6) (7) (8)

Loadings for F1
  X1           1 2 3 4 5 6 7 8
  X2           1 2 3 4 5 6 7 8
  X3           1 2 3 4 5 (6) (7) (8)
  X4           1 2 3 4 5 6 7 8
  X5           1 2 3 4 5 6 7 8
  X6           1 2 3 4 5 6 7 8
  X7           1 2 3 4 5 (6) (7) (8)

Loadings for F2
  X8           1 2 3 4 5 6 7 8
  X9           1 2 3 4 5 6 7 8
  X10          1 2 3 4 5 6 7 8
  X11          1 2 3 4 5 6 7 8
  X12          1 2 3 4 5 6 7 8
  X13          1 2 3 4 5 (6) (7) (8)
  X14          1 2 3 4 5 6 7 8
```

Results MMG-FA (Noninvariant)

	nr of clusters	loglik	nrpars	BIC_N	BIC_G	screeratios	convergence
[1,]	1	-85948.70	176	173357.2	172263.4	NA	1
[2,]	2	-85658.67	201	172984.4	171735.3	2.384504	1
[3,]	3	-85537.04	226	172948.5	171544.0	1.565931	1
[4,]	4	-85459.37	251	173000.5	171440.7	5.143030	1
[5,]	5	-85444.26	276	173177.7	171462.5	1.046035	1
[6,]	6	-85429.82	301	173356.2	171485.6	NA	1



Scree ratio is max at 2 cluster solution. Elbow at 3 clusters. 3 cluster solution nearly correct.

	Cluster_1	Cluster_2
1	0	1
2	0	1
3	1	0
4	0	1
5	0	1
6	0	1
7	0	1
8	0	1

	Cluster_1	Cluster_2	Cluster_3
1	0	0	1
2	0	0	1
3	0	1	0
4	0	0	1
5	0	0	1
6	1	0	0
7	1	0	0
8	1	0	0

Results MMG-FA (Noninvariant)

Cluster 1

	Factor_1	Factor_2
x1	1.0369532	0.0000000
x2	0.6088257	0.0000000
x3	0.3517281	0.0000000
x4	0.4799559	0.0000000
x5	0.9026586	0.0000000
x6	0.8027759	0.0000000
x7	0.2480157	0.0000000
x8	0.0000000	1.0015064
x9	0.0000000	0.4802230
x10	0.0000000	0.7449974
x11	0.0000000	0.8426073
x12	0.0000000	0.7298424
x13	0.0000000	0.2588983
x14	0.0000000	0.7248126

Cluster 2

	Factor_1	Factor_2
x1	0.9740124	0.0000000
x2	0.5517547	0.0000000
x3	0.7418591	0.0000000
x4	0.4323821	0.0000000
x5	0.9635197	0.0000000
x6	0.6868613	0.0000000
x7	0.5924956	0.0000000
x8	0.0000000	1.0122017
x9	0.0000000	0.4485863
x10	0.0000000	0.8339799
x11	0.0000000	0.8880168
x12	0.0000000	0.6232514
x13	0.0000000	0.5944232
x14	0.0000000	0.7331219

Cluster 3

	Factor_1	Factor_2
x1	1.0324087	0.0000000
x2	0.6102738	0.0000000
x3	0.6831400	0.0000000
x4	0.5456828	0.0000000
x5	0.9480776	0.0000000
x6	0.8213167	0.0000000
x7	0.6458287	0.0000000
x8	0.0000000	0.9771686
x9	0.0000000	0.4709385
x10	0.0000000	0.7679590
x11	0.0000000	0.8722622
x12	0.0000000	0.7099456
x13	0.0000000	0.5621034
x14	0.0000000	0.7260020

	x1	x2	x3	x4	x5	x6	x7	x8
Cluster_1	0.02233440	-0.05573626	0.007298989	-0.04779280	0.0074936387	-0.04221381	-0.04131316	-0.04580363
Cluster_2	0.02996339	0.47373371	0.603915674	0.51700338	0.0037079379	-0.05787138	-0.05820922	-0.13595281
Cluster_3	0.06025034	-0.01737046	-0.007499078	-0.06076466	0.0004349774	-0.03590530	0.02564286	-0.07981347

	x9	x10	x11	x12	x13	x14
Cluster_1	-0.02331586	-0.06710739	0.01181938	0.12976122	0.11626914	0.11049842
Cluster_2	0.02011012	0.48128418	0.64196924	0.46528166	-0.07203630	-0.06958884
Cluster_3	0.01568385	-0.04894289	-0.01508219	-0.07508553	-0.01117446	-0.04488238

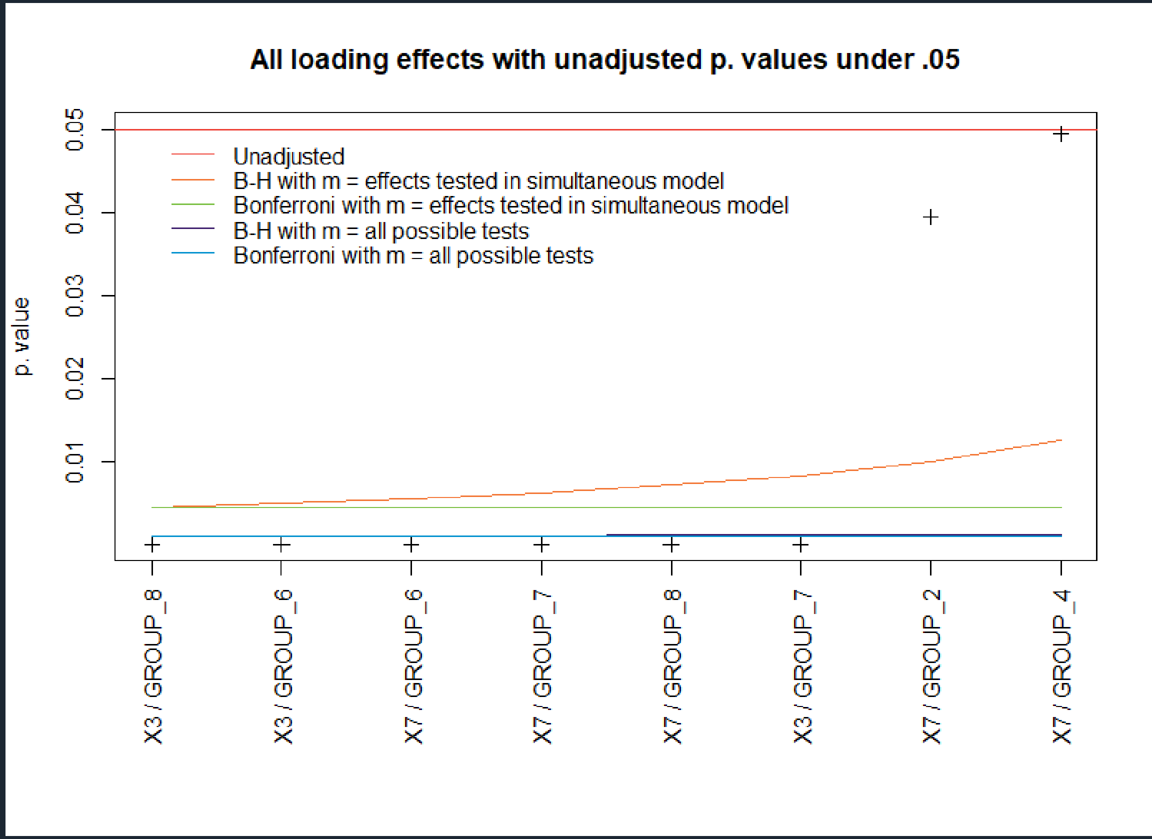
Recall Noninvariant Model

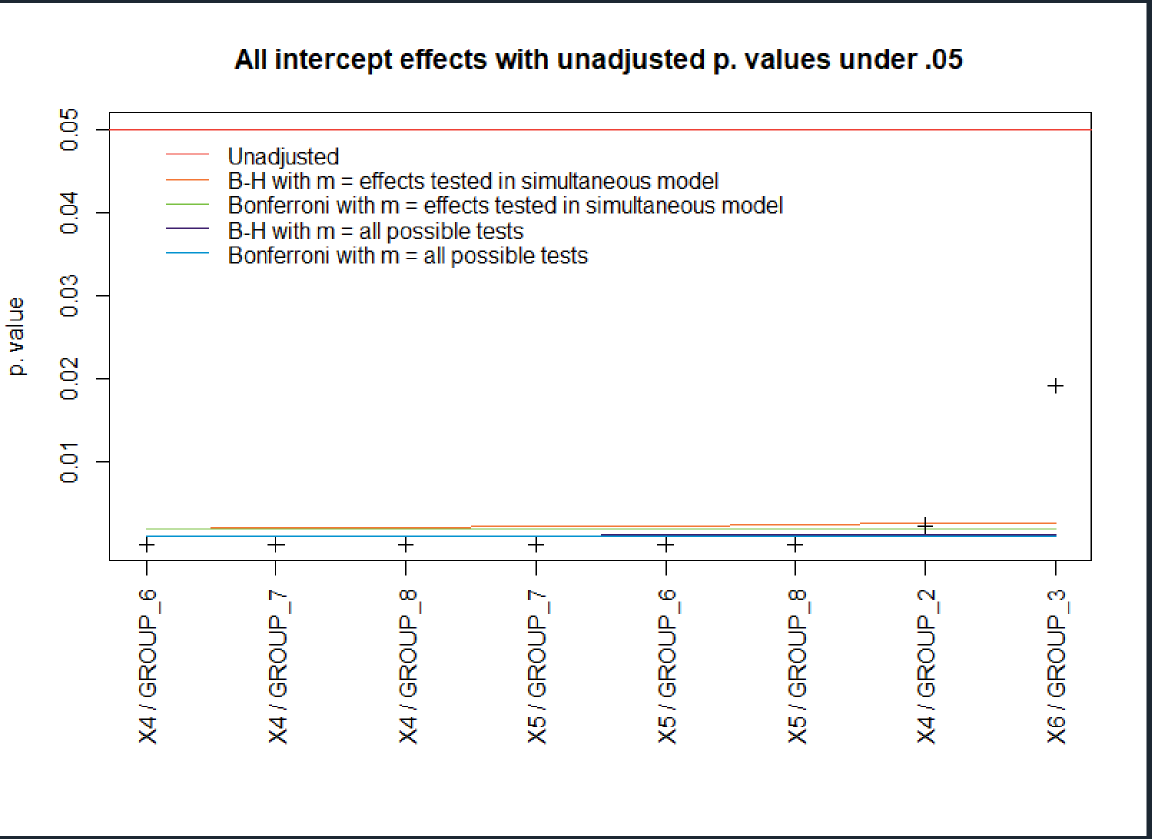
- Noninvariant model
  - medium-large factor loading noninvariance for items 3, 7, and 13 for groups 6-8
  - Medium item intercept noninvariance for items 4, 5, and 14 for groups 6-8
    - Constant of 0.5 to items

```
noninvariant_model <- 'f1=~x1+.6*x2+.35*x3+.5*x4 + 0.9*x5 + 0.8*x6 + 0.25*x7
                       f2=~x8+.5*x9+.8*x10+.9*x11 + 0.7*x12 + 0.25*x13 + 0.7*x14
                       f1 ~~ 0.35*f2'
```

Results: MNLFA (Noninvariant)

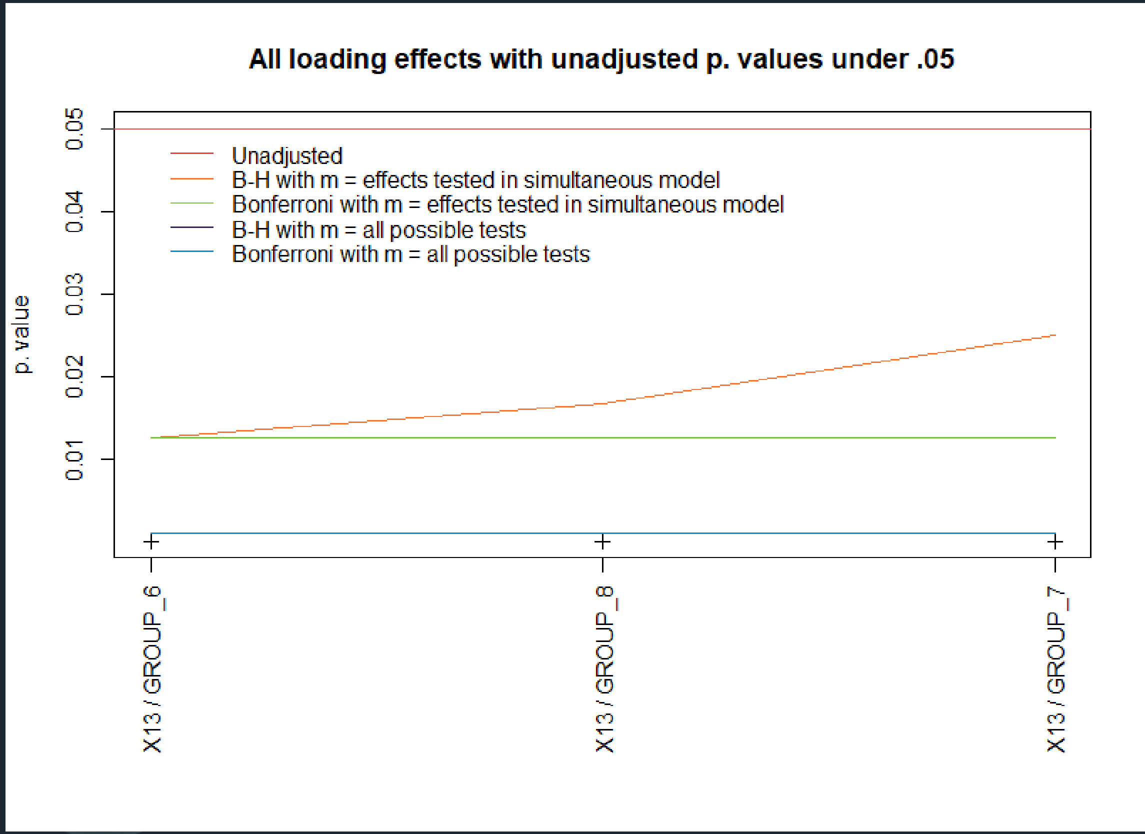
Factor 1

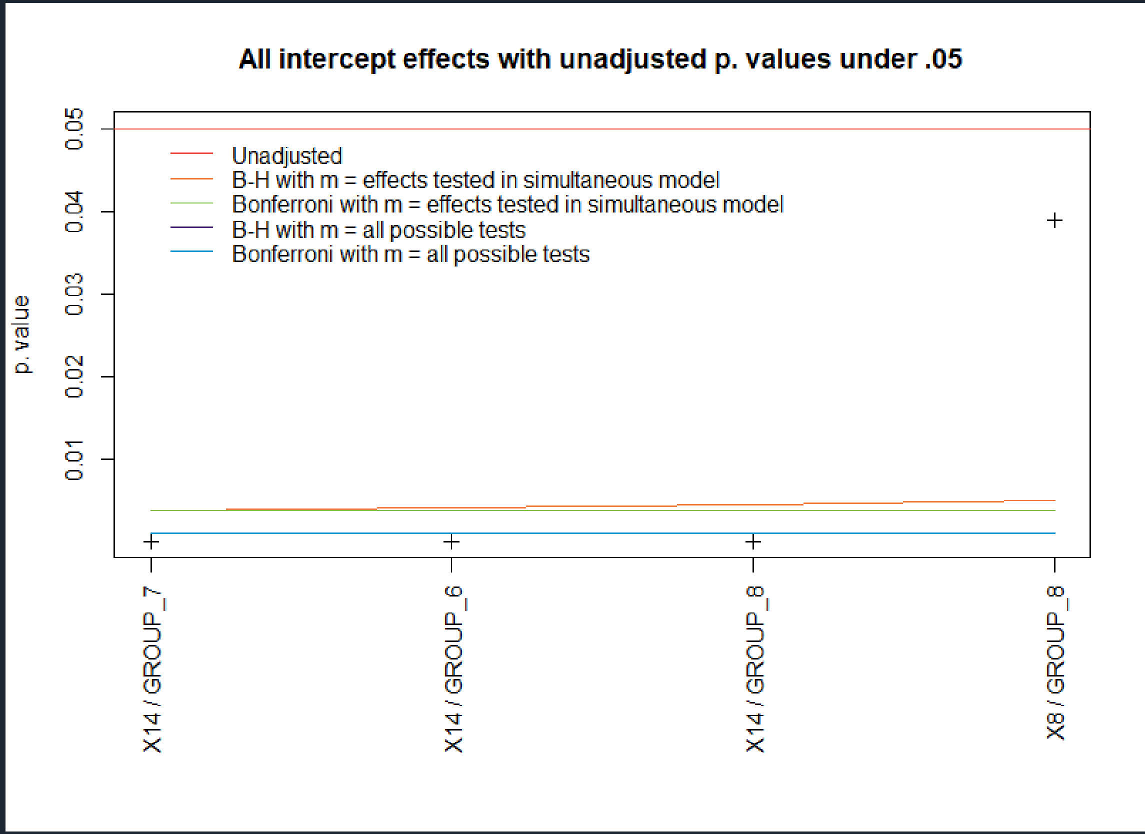




Results: MNLFA (Noninvariant)

Factor 2





Empirical Demonstration

Data

- Analyze the eight-item Patient Health Questionnaire depression instrument (PHQ-8) in a sample of 30,215 American adults in the United States from the 2019 National Health Interview Survey (NHIS)
- Examine the MI of the PHQ-8 depression instrument across 16 intersectional subgroups defined at the intersection of
  - age (under 52, 52 and older),
  - gender (male, female),
  - race (Black, non-Black), and
  - education (no bachelor's degree, with bachelor's degree)

8-item Patient Health Questionnaire (PHQ)

Table 1

Patient Health Questionnaire (PHQ) depression instrument items

Item	Item Text
	Over the last two weeks, how often have you been bothered by ...
PHQ-1	Little interest or pleasure in doing things?
PHQ-2	Feeling down, depressed, or hopeless?
PHQ-3	Trouble falling or staying asleep, or sleeping too much?
PHQ-4	Feeling tired or having little energy?
PHQ-5	Poor appetite or overeating?
PHQ-6	Feeling bad about yourself, or that you are a failure, or have let yourself or your family down?
PHQ-7	Trouble concentrating on things, such as reading the newspaper or watching television?
PHQ-8	Moving or speaking so slowly that other people could have noticed? Or the opposite, being so fidgety or restless that you have been moving around a lot more than usual?

Note. The response options for each item were not at all (1), several days (2), more than half the days (3), or nearly every day (4).

Intersectional Subgroups

- 16 intersectional groups
- 52+ Black males with no college degree (smallest subgroup)
- 52+ Non-Black females with no college degree (largest subgroup)
- Internal consistency (alpha) relatively high across all groups

Table 1

Intersectional group definitions and internal consistencies on PHQ-8

Group Code	Group	N	Internal consistency
1	< 52 Black Females with College Degree	464	0.84
2	< 52 Black Females with No College Degree	639	0.87
3	< 52 Black Males with College Degree	294	0.86
4	< 52 Black Males with No College Degree	472	0.86
5	< 52 Non-Black Females with College Degree	3654	0.84
6	< 52 Non-Black Females with No College Degree	2679	0.88
7	< 52 Non-Black Males with College Degree	3119	0.84
8	< 52 Non-Black Males with No College Degree	2817	0.87
9	52+ Black Females with College Degree	380	0.84
10	52+ Black Females with No College Degree	595	0.84
11	52+ Black Males with College Degree	498	0.84
12	52+ Black Males with No College Degree	209	0.87
13	52+ Non-Black Females with College Degree	3717	0.84
14	52+ Non-Black Females with No College Degree	4195	0.84
15	52+ Non-Black Males with College Degree	3199	0.83
16	52+ Non-Black Males with No College Degree	3284	0.84
	Overall	30215	0.85

Note. Internal consistencies measured using Cronbach's alpha.

Note: Traditional MI Testing

- With this data, we were able to establish configural, metric, and salar invariance across all 16 groups using a traditional multiple-group CFA approach for MI
  - May not always be the case with many groups
  - Internal consistency for the PHQ-8 was very high across the groups in this sample

```
################### Nested Model Comparison #########################

Chi-Squared Difference Test

                 Df    AIC    BIC  Chisq Chisq diff    RMSEA Df diff Pr(>Chisq)
fit_configural  320 577012 580194 7487.3
fit_metric      425 577877 580190 8563.2    1075.86 0.070959     105  < 2.2e-16 ***
fit_scalar      530 578553 579995 9448.9     885.71 0.063632     105  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

####################### Model Fit Indices ###########################
                  chisq  df pvalue rmsea   cfi   tli  srmr        aic         bic
fit_configural 7487.307† 320   .000 .110  .917† .884  .044† 577011.510† 580194.142
fit_metric     8563.163  425   .000 .102  .906  .901  .062  577877.366  580189.747
fit_scalar     9448.872  530   .000 .096† .897  .913† .065  578553.075  579995.205†

################## Differences in Fit Indices #######################
                               df  rmsea    cfi   tli  srmr     aic      bic
fit_metric - fit_configural 105 -0.008 -0.011 0.017 0.018 865.855   -4.395
fit_scalar - fit_metric     105 -0.006 -0.009 0.012 0.003 675.709 -194.542
```

Results: AM

- The table notes which item thresholds and loadings are non-invariant in which groups
- The results indicate that, even after alignment, there are many item parameters that remain noninvariant in several of the groups
- Overall, we can see that only 5% of the thresholds (6 out of 128) are non-invariant and 24% of the factor loadings are non-invariant (31 out of 128)
  - Using the 25% rule-of-thumb presented by Muthén and Asparouhov (2014), the results imply trustworthy alignment results

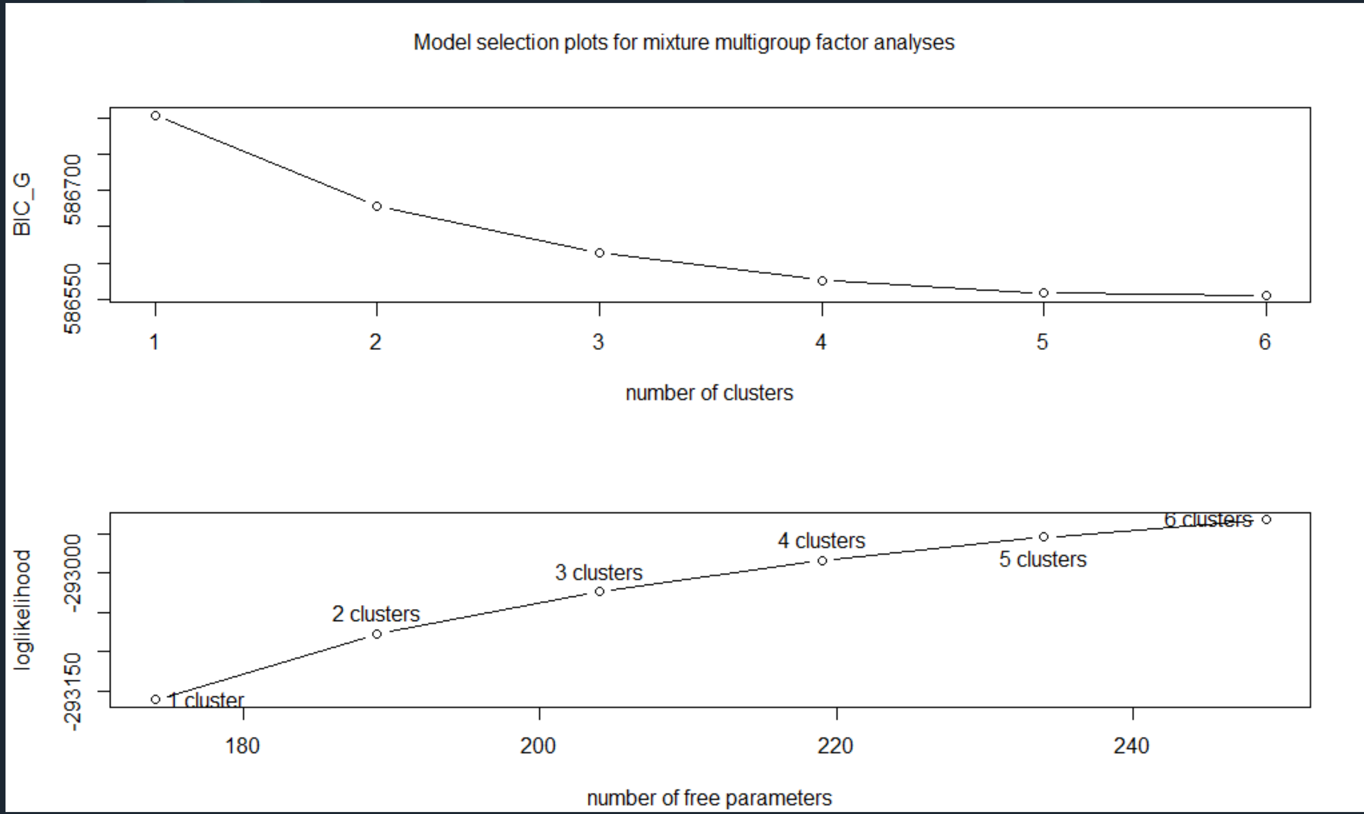
TABLE 2 Invariance results for aligned intercept and loading parameters for PHQ-1 to PHQ-8.

Item intercepts			
Item	**Group invariance (non-invariance)**	**# Non-invariant groups**	**Invariance index**
PHQ-1	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	0	0.750
PHQ-2	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	0	0.848
PHQ-3	1 2 3 4 5 6 7 8 9 10 11 12 **(13)** **(14)** **(15)** 16	3	0.662
PHQ-4	1 2 3 4 **(5)** **(6)** 7 8 9 10 11 12 13 **(14)** 15 16	3	0.670
PHQ-5	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	0	0.818
PHQ-6	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	0	0.608
PHQ-7	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	0	0.841
PHQ-8	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	0	0.636
Factor loadings			
Item	**Group (non)-invariance**	**# Non-invariant groups**	**Invariance index**
PHQ-1	**(1)** **(2)** 3 4 **(5)** **(6)** 7 **(8)** 9 10 11 12 13 14 15 16	5	0.501
PHQ-2	**(1)** **(2)** 3 4 **(5)** **(6)** 7 8 9 10 11 12 13 14 15 16	4	0.594
PHQ-3	1 2 3 4 **(5)** 6 7 8 9 10 11 12 13 **(14)** 15 16	2	0.809
PHQ-4	1 2 3 4 5 **(6)** 7 8 9 10 11 12 **(13)** 14 15 16	2	0.681
PHQ-5	1 2 3 **(4)** 5 6 **(7)** **(8)** 9 10 11 12 13 **(14)** **(15)** **(16)**	6	0.632
PHQ-6	1 2 3 4 **(5)** **(6)** **(7)** **(8)** 9 **(10)** 11 12 13 14 15 **(16)**	6	0.590
PHQ-7	1 2 3 4 5 6 7 8 9 10 11 12 **(13)** 14 15 16	1	0.884
PHQ-8	**(1)** 2 3 4 **(5)** 6 7 8 9 10 11 12 **(13)** **(14)** 15 **(16)**	5	0.341

Note: The group values correspond to the intersectional coding (see Table 1). The bolded numbers in parentheses refer to the groups that show significant non-invariance for the parameter. The invariance index is R^2. An R^2. close to 1 provides evidence that there is complete invariance. Conversely, an R^2. near 0 provides evidence that group mean differences explain little to no variability in item parameters.

Cintron et al. (in press)

Results: MMG-FA Clusters with Scalar Invariance



- The methods to determine the number of clusters indicated that a 1-cluster solution may be likely (i.e., there is evidence of measurement invariance across the 16 intersectional subgroups)
  - There was not clear evidence of an elbow suggesting 1 cluster may be optimal (see De Roover, 2014)
  - Scree ratio largest at 2 clusters
- Examine the two-cluster solution here for demonstration
  - As we will see, the two-cluster solution only captures subtle intercept differences
  - In this scenario, De Roover (2014) suggested that the more parsimonious solution is preferable (here 1-cluster)

Results: MMG-FA Clusters with Scalar Invariance

- Methods for selecting the number of clusters suggested a two-cluster solution (see De Roover, 2001)
- Cluster 1 (38%)
  - 52+ Non-Black Males without a college degree,
  - <52 Non-Black Males with a college degree,
  - 52+ Non-Black Females with a college degree,
  - <52 Non-Black Females without a college degree,
  - <52 Black Males with a college degree
  - <52 Black Females with a college degree
- Cluster 2 (62%)
  - <52 Non-Black Males without college degree,
  - 52+ Non-Black Males with a college degree,
  - <52 Non-Black Females with a college degree,
  - 52+ Non-Black Females without a college degree,
  - <52 Black Males without a college degree,
  - 52+ Black Females with a college degree,
  - 52+ Black Females without a college degree,
  - <52 Black Females without a college degree,
  - 52+ Black Males with a college degree,
  - 52+ Black Males without a college degree

	Cluster 1	Cluster 2	Cluster 1	Cluster 2
	Loadings		Intercepts	
Item 1	0.74	0.74	0.07	-0.00
Item 2	0.78	0.78	0.05	-0.03
Item 3	0.59	0.59	0.02	-0.00
Item 4	0.65	0.64	0.02	-0.04
Item 5	0.63	0.62	-0.02	-0.00
Item 6	0.71	0.72	-0.00	0.02
Item 7	0.63	0.65	-0.00	0.01
Item 8	0.53	0.54	0.01	-0.01

Results: MNLFA

- Explored measurement noninvariance for factor loadings and item intercepts
- Experienced convergence issues
- Computationally intensive to run relative to other methods

Discussion

- The theory of intersectionality has provided new opportunities for the quantitative analysis of data
  - Multilevel analysis of individual heterogeneity and discrimination accuracy (MAIHDA; Evans et al., 2018)
  - Differential item functioning (DIF) analyses using an intersectional lens (Russell & Kaplan, 2021)
- We provided an example of using intersectionality theory to inform measurement invariance testing
- We found multiple forms of evidence of *intersectional measurement invariance* of the PHQ-8 instrument across the 16 intersectional subgroups

Limitations

- Not exploring all possible intersections
  - Some intersections resulted in very small N and were not possible to evaluate
    - Only 10 Non-Hispanic Black Males with a college degree in the 30,000+ NHIS sample
  - We performed an intercategorical intersectional analysis as opposed to an intracategorical intersectional analysis (Bauer & Scheim, 2019)
  - Choice of intersections to evaluate should be driven by theoretical considerations
- Small N in some groups
  - Traditional MI testing research suggests that each group should have a sample size of at least 400 (French & Finch, 2006)
- Treated Likert-scaled items as continuous
  - Treating items as categorical (i.e., ordinal) may result in more valid inference
  - *mixmgfa* not able to handle categorical responses

Conclusions

- We need to consider whether response behavior on an instrument may vary as a function of the intersection of individuals' multiple identities
- The alignment method, mixture multiple-group factor analysis, and moderated nonlinear factor analysis are each ways to perform intersectional measurement invariance testing
  - Alignment method has benchmarks and Monte Carlo simulation evidence supports its use to validate an instrument (i.e., includes the 25% benchmark for determining whether results of scale are trustworthy across groups)
  - Mixture multiple-group factor analysis unique in its ability to find clusters of groups with invariance
  - Moderated nonlinear factor analysis was computationally intensive, but may be useful when the focus is on different item functioning
    - Flexibility in considering where different forms of noninvariance emerge across intersectional identities (e.g., in the main effects or at certain intersections, or margins, of an identities)
    - Challenging with this large-scale dataset (Bayesian estimation may be useful)
- Future research will include examining the goodness of recovery of the group factor means, factor loadings, and item intercepts across the three methods

References

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. Structural Equation Modeling: A Multidisciplinary Journal, 21(4), 495-508.

Bauer, G. R., & Scheim, A. I. (2019). Advancing quantitative intersectionality research methods: Intracategorical and intercategorical approaches to shared and differential constructs. Social Science & Medicine, 226, 260-262

Bešić, E. (2020). Intersectionality: A pathway towards inclusive education?. Prospects, 49(3-4), 111-122.

Cintron, D. W., Matthay, E. C., & McCoach, D. B. (2023). Testing for intersectional measurement invariance with the alignment method: Evaluation of the 8-item patient health questionnaire. Health Services Research.

Curran, P. J., McGinley, J. S., Bauer, D. J., Hussong, A. M., Burns, A., Chassin, L., ... & Zucker, R. (2014). A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. Multivariate behavioral research, 49(3), 214-231.

De Roover, K. (2021). Finding clusters of groups with measurement invariance: Unraveling intercept non-invariance with mixture multigroup factor analysis. Structural Equation Modeling: A Multidisciplinary Journal, 28(5), 663-683.

Luong, R., & Flake, J. K. (2022). Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting. Psychological Methods.

Muthén, B., & Asparouhov, T. (2014). IRT studies of many groups: The alignment method. Frontiers in psychology, 5, 978.

Thank you!

dwc237@cornell.edu

cintrond.github.io

@cintrondw