An Introduction to Measurement Invariance Testing: Resource Packet for Participants

Do our Measures Measure up? The Critical Role of Measurement Invariance

Demonstration Session
American Evaluation Association, October 2013
Washington, DC

Silvana Bialosiewicz, MA silvana@cgu.edu
Kelly Murphy, MA kelly.murphy@cgu.edu
Tiffany Berry, PhD tiffany.berry@cgu.edu

Claremont Evaluation Center
School of Social Science, Policy, & Evaluation
Claremont Graduate University
123 East Eighth Street
Claremont, CA 91711

Table of Contents

Section One: ***An Introduction to Measurement Invariance*** …………………………………… **3**

Section Two: ***Annotated Mplus Output*** ………………………………………………………….. **10**

Section Three: ***Software Options*** …………………………………………………………… **35**

Section Four: ***Additional Resources*** ………………………………………………………….. **36**

Section 1: *Introduction to Measurement Invariance*

Definition

Measurement Invariance: The statistical property of a measurement that indicates that the same underlying construct is being measured across groups or across time.

Also referred to as **factor invariance, factorial invariance, factor equivalence**

How do we know when we have it?
When the relationship between manifest indicator variables (scale items, subscales, etc.) and the underlying construct are the same across groups or across time.

Two Types of Measurement Invariance Tests:

1. Multi-group invariance: Does the model hold across groups (e.g., males and females, child and adult participants).
2. Longitudinal Invariance: Does the model hold across time (e.g., pre and post test).

Measurement Invariance and Program Evaluation

The outcomes we explore in our evaluations are often complex, multi-dimensional constructs that cannot be directly observed such as participant attitudes and beliefs, intentions and motives, and emotional and mental states. In program evaluation, self-report surveys are one of the most common methods of exploring the relationship between these constructs and participation in the program.

Within our field, there is a large amount of variability in how these surveys are developed:

- ✓ When we are using previously validated scales, we are often using them on populations that are quite different from the one in which the scale was validated
- ✓ It is common for us to begin with validated scales but then make changes to fit the evaluation context and participant populations.
- ✓ It is often the case that no measure exists that maps on to our evaluation question and measures must be developed specifically for the project.

In all of these scenarios our ability to assess true differences between groups or change over time can be hindered by measurement error.

Measurement error can affect our ability to make accurate and meaningful comparisons between groups or across time points when determining the impact of a program. So, in addition to traditional tests of reliability and validity, we can perform tests of measurement invariance to answer these important questions such as these: [1]

- Do different groups of respondents interpret a given measure in a conceptually similar manner?
- Does participation in the program alter the conceptual frame of reference against which a group responds to a measure over time?

Answering questions such as these in a statistically rigorous manner helps us ensure that the comparisons we make represent true differences in our constructs of interest.

When to conduct tests of measurement invariance[2]

- ✓ When evaluation involves comparisons, between individuals or groups, and differences are assumed to have substantive meaning
- ✓ When comparisons involve data collected from a self-report survey
- ✓ When the survey is comprised of one or more sets of items, which when combined are intended to assess a construct or constructs
- ✓ When there is evidence of the instrument's psychometric quality (i.e., tests of reliability and validity) and that the common factor model holds with the data (i.e., confirmatory factor analysis)

Introduction to the CFA framework

Often, our constructs of interest are measured using multi-item scales. One example would be an adapted version of Rosenberg's self-esteem scale using the items, "I am able to do things as well as most other people", "I feel that I have a number of good qualities", and "I take a positive attitude toward myself". Each item on its own is insufficient to capture the construct of interest, but together, we hope the represent a valid indirect assessment of the construct. We combine these items to form a composite measure, with the assumption that the composite will be a more reliable estimate of the construct than any one item on its own. Because the construct cannot be directly observed, it is referred to as a latent variable or a common factor. The confirmatory factor analysis (CFA) framework provides a means to test the construct validity of an item set as an indirect measure of this hypothesized latent variable.

[1] (Vandenberg & Lance, 2000)
[2] (Vandenberg & Lance, 2000)

Figure 1 illustrates a simple common factor model with one latent variable measured by three indicator items.

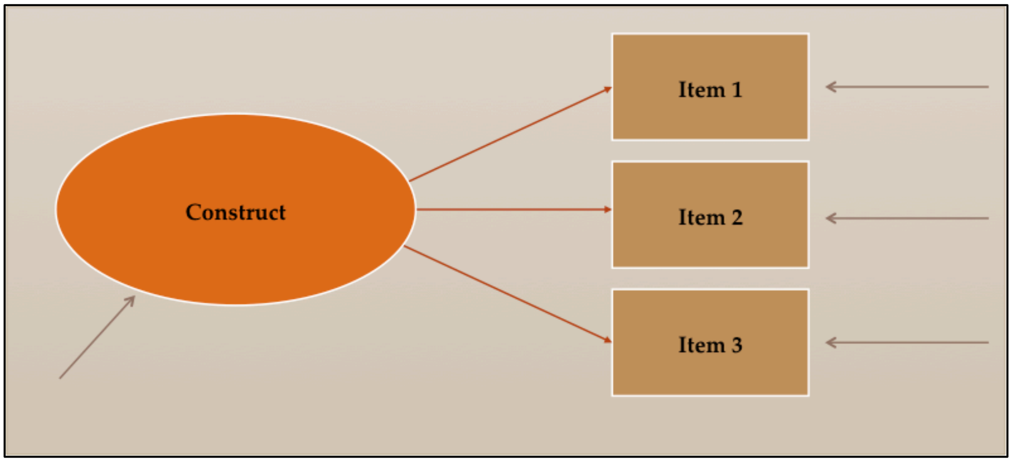

Figure 1. A Common Factor Model

Within this model are a number of components with which you should be familiar:

- ✓ Common Factor: The latent variable that represents a theoretical construct that cannot be directly observed
- ✓ Observed/manifest variables: The tangible measures (i.e., survey items or item sets) that serve as indicators for the latent variable
- ✓ Factor loadings: The correlation between items and construct (how highly the item "loads" onto the latent factor). These are structural regression coefficients, which represent the magnitude of expected change in the observed variables for every change in the latent variable. The unidirectional arrows between the construct and the observed variables represent the factor loadings
- ✓ Item intercepts: The origin or starting value of the scale that your factor is based on. These are the same as a regression intercept. These values are not visually represented in the model.
- ✓ Latent factor mean: The construct mean, which we are attempting to measure by our indicator variables. This value is not visually represented in the model.
- ✓ Factor variance: Known as residual error, this represents the overall error in prediction of your construct using our indicator variables. The unidirectional arrow pointing at our latent variable represents this variance.

- ✓ Variances and Covariances: The measurement error associated with each observed variable. The unidirectional arrows pointing at the observed variables represent these error terms.

Data = Model + Residual

Our statistical tests explore the fit between the theorized model and the data collected from participants. So when we're conducting a confirmatory factor analysis we are testing the match between the data and the hypothesized model of one or more latent variables. In very simple terms, we force our data into our hypothesized model, and then see how well it fits. What we have then is variance explained by our model and the residual, the variance that cannot be explained by the model.

Model Fit Indices[3]

When we are assessing model fit at each stage of the measurement invariance testing process, we use the fit indices described in Figure 2. The field has not reached absolute agreement on what constitutes a good fit, but there is general consensus in the values presented here, although some advocate for even stricter standards (please refer to the articles in Section Four of this packet for more in depth discussions of fit indices standards).

Chi Square	• Smaller is better • Sensitive to sample size
CFI/TLI	• Acceptable >.90 • Good fit >.95
RMSEA	• Unacceptable >.10 • Good Fit <.06
SRMR	• Good fit < .05

Figure 2. Model Fit Indices for Measurement Invariance Testing

[3] (Byrne, 2012)

Chi-Square: In this context the chi-squared value is the likelihood-ratio test statistic. The chi-squared tests the differences between the observed data and model covariance matrix. Our goal is *not* to reject the null hypothesis (that the two are significantly different) and when we fail to reject the null that is indication of good fit. So what we're looking for non-significant p-values and a small chi-square. However, the chi-square statistic is very sensitive to sample size so its easy to get significance with large samples. At each stage of measurement invariance testing, we use a chi-square different test as our primary indication of model fit (e.g., whether we have attained a new level of measurement invariance).

Due to the chi-square's limitations, researchers have developed a few other fit indices which we use to evaluate the fit of our model The 2 most commonly used are the **CFI** and the **TLI.** These measure the improvement in model fit from a non-restricted model to the hypothesized restricted model. The CFI ranges from 0-1 and the TLI can exceed 1. For both, 0.90 or above is seen as acceptable and 0.95 or above is considered a good fit

The **RMSEA** & the **SRMR** are known as absolute "misfit" indices. These indices tell us how well the hypothesized model fits the sample data. These indices *decrease* as fit improves therefore the lower the value the better. Generally agreed that a RMSEA, less than .06 is a good fit and those around .8 are acceptable. The SRMR ranges from 0-1 with indication of a good fit for values less than 0.05.

Levels of Measurement Invariance

As shown in Figure 3, there are essentially four levels of measurement invariance and each of these levels builds upon the previous by introducing additional equality constraints on model parameters to achieve stronger forms of invariance. As each set of new parameters is tested, the parameters known to be invariant from previous levels are constrained. Thus, the process of assessing measurement invariance is essentially the testing of a series of increasingly restrictive hypotheses.

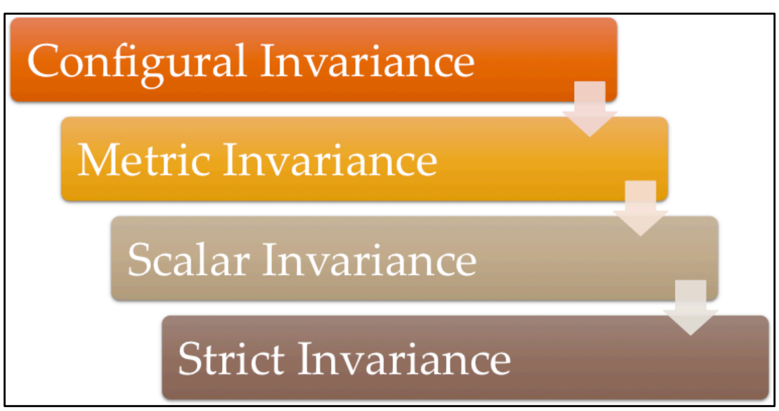

Figure 3. Levels of Measurement Invariance

Configural Invariance: When assessing measurement invariance, you begin with the establishment of configural invariance. In the measurement invariance literature configural invariance is also commonly referred to as pattern invariance and is considered to be the baseline model. In this level we are only interested in testing whether or not the same items measure our construct across administrations (e.g. across multiple groups or across time). To test this, we estimate both factor models simultaneously. Because this is the baseline model you only need to assess overall model fit to test whether configural invariance holds

Metric Invariance: This level of invariance is also commonly referred to as weak invariance. Metric invariance builds upon configural invariance by requiring that in addition to the constructs being measured by the same items, the factor loadings of those items must be equivalent across administrations. Factor factor loadings reflect the degree to which differences among participants' responses to the item arise from differences among their levels of the underlying construct that is being assessed by that item. Thus, attaining invariance of factor loadings suggests that the construct has the same meaning to participants across administrations. The reasons this is the case is because if a construct has the same meaning across administrations then we would expect identical relationships between the construct and the participants responses to the items used to measure the construct.

To assess metric invariance we compare the fit of the metric model with the fit of the configural model using a chi-square difference test. If there is no significant difference in model fit than there is evidence to suggest that the factor loadings are invariant across administrations. Attaining metric invariance suggests that group comparisons of factor variances and covariances are defensible. However, it does not justify the comparisons of group means.

Scalar Invariance: The ability to justify mean comparisons across time or across groups is established by attaining scalar or strong invariance. Scalar invariance builds upon metric invariance by requiring that the item intercepts also be equivalent across administrations. Item intercepts are considered the origin or starting value of the scale that your factor is based on. Thus, participants who have the same value on the latent construct should have equal values on the items the construct is based.

To assess scalar invariance we compare the fit of the scalar model with the fit of the metric model. If there is no significant difference in model fit than there is evidence to suggest intercept invariance. Non-invariance of intercepts may be indicative of potential measurement bias and suggests that there are larger forces such as cultural norms or developmental differences that are influencing the way that participants are responding to items across administrations and that participants are systematically rating items either higher or lower at each administration time.

Strict Invariance: The final level of invariance is called strict factorial invariance. Unlike the previously discussed levels of measurement invariance, there are two sublevels of strict invariance. The first level of strict invariance is invariance of factor variances. Factor variances represent the overall error in the prediction of your construct. The second level of strict invariance refers to invariance of individual indicator variable's error terms that represent the unique error specific to that particular indicator variable. So when testing strict invariance you are essentially testing whether your residual error is equivalent across administrations.

Because there are two levels of strict invariance, you assess strict invariance across two models. First, you estimate the model with the constrained factor variances. After establishing invariance of factor variances, you estimate the model with constrained error variances. Similar to the previous levels of measurement invariance, strict invariance is tested through a chi-square difference test with the preceding model. It should be noted, however, that strict invariance represents a highly constrained model and is rarely achieved in practice. Because of this most experts in the field now agree that it's too unreasonable to expect equality in residual variances across groups or across time.

Section 2: *Annotated Mplus Output*

Below you will find annotated syntax and output for each level of measurement invariance. Colored font represents the standard commands used in Mplus syntax. Regular font represents the variable commands specific to this dataset. **Bolded text** represents descriptions of syntax and output.

Configural Model for Longitudinal Invariance: Input
TITLE: CONFIGURAL MODEL SE MS –**Title you gave your analysis.**
DATA: file is MSsurvey.dat; – **Save your input in the same folder as your data and enter name of dataset here.**
VARIABLE: – **This is the start of the command where you describe your variables.**
NAMES ARE ID GENDER GRADE SE1 SE2 SE3 SE4 SE1P SE2P SE3P SE4P;– **All variables must be listed in the SAME order as listed in database.**
USEVARIABLES ARE SE1-SE4 SE1P-SE4P; – **List only variables used in current analysis.**
MISSING=ALL (999); – **This is how you tell Mplus what your missing data code is.**
MODEL: – **Enter model commands under here.**
SEPRE BY SE1– **You can name your latent variable anything. The BY command tells Mplus**
SE2 **what indicator variables your latent variable is measured by.**
SE3
SE4;
SEPOST BY SE1P
SE2P
SE3P
SE4P;
[SEPRE SEPOST]; – **You are requesting latent variable means with this command.**
[SE1@0 SE1P@0]; – **You are constraining the intercepts of the reference variable to be 0.**
OUTPUT: MODINDICES(ALL 0); – **Here we are requesting all different types of modification indices regardless of how small the estimated chi-square change will be.**

Configural Model for Longitudinal Invariance: Selected output

Chi-Square Test of Model Fit – **Here we have our chi-square value and degrees of freedom we use to conduct our chi square difference tests.**

Value	36.569
Degrees of Freedom	19
P-Value	0.0090

MODEL RESULTS – **Below we have our unstandardized results. You will want to check these estimates to make sure all parameters were estimated properly and that parameter estimates are in the expected direction.**

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value	
SEPRE BY					**– These are the factor loading estimates**
SE1	1.000	0.000	999.000	999.000	**– The reference variable does not have an estimate due to requirements for latent variable scaling.**
SE2	0.988	0.044	22.582	0.000	
SE3	0.964	0.045	21.341	0.000	
SE4	1.035	0.045	22.813	0.000	
SEPOST BY					
SE1P	1.000	0.000	999.000	999.000	
SE2P	0.996	0.043	23.120	0.000	
SE3P	0.996	0.046	21.755	0.000	
SE4P	1.013	0.046	22.039	0.000	
SEPOST WITH					
SEPRE	0.322	0.034	9.489	0.000	**This is the covariance of self-efficacy at pre-test and post-test.**
Means					
SEPRE	3.633	0.027	132.453	0.000	**–Estimates of the latent means at pre-test and post-test.**
SEPOST	3.515	0.031	113.416	0.000	
Intercepts					**–Here are our estimates for our item intercepts.**
SE1	0.000	0.000	999.000	999.000	**–The reference variables was constrained to enable Mplus to**

SE2	-0.073	0.162	-0.450	0.653 **estimate the factor means.**
SE3	-0.033	0.168	-0.196	0.845
SE4	0.198	0.168	1.178	0.239
SE1P	0.000	0.000	999.000	999.000
SE2P	-0.146	0.155	-0.940	0.347
SE3P	-0.201	0.164	-1.221	0.222
SE4P	0.232	0.165	1.409	0.159

Variances –**Here are the estimates for our factor variances.**

SEPRE	0.738	0.050	14.738	0.000
SEPOST	0.804	0.056	14.266	0.000

Residual Variances –**Here are the estimates for our error variances.**

SE1	0.987	0.040	24.774	0.000
SE2	0.988	0.040	25.012	0.000
SE3	1.126	0.043	26.456	0.000
SE4	0.901	0.039	23.235	0.000
SE1P	0.878	0.041	21.485	0.000
SE2P	0.816	0.039	21.077	0.000
SE3P	0.984	0.044	22.412	0.000
SE4P	0.887	0.041	21.402	0.000

MODEL MODIFICATION INDICES–**Modification indices inform us of badly chosen parameter constraints.**

Minimum M.I. value for printing the modification index 10.000

	M.I.	E.P.C.	Std E.P.C.	StdYX E.P.C.
WITH Statements				
SE5P WITH SE4P	12.930	0.132	0.132	0.141

–**Our chi-square will drop by approximately 13 if we correlate the errors of these two items. We will ignore this recommendation because we already have a good fitting model and we want to keep our model as parsimonious as possible.**

Metric Model for Longitudinal Invariance: Selected Input

```
MODEL:
  SEPRE BY SE1
              SE2 (1)  – Add parentheses next to corresponding factor loadings to constrain them to be
              SE3 (2)    equal.
              SE4 (3);
  SEPOST BY SE1P
               SE2P (1)
               SE3P (2)
               SE4P (3);
  [SEPRE SEPOST];
   [SE1@0 SE1P@0];
OUTPUT: MODINDICES(ALL 0);
```

Metric Model for Longitudinal Invariance: Selected Output

```
Chi-Square Test of Model Fit – Chi-square and dfs are used to conduct chi-square difference test with
     Value                    37.345  configural model.
     Degrees of Freedom         22 – We gained 3 df because we constrained three factor loadings to be
      P-Value                  0.0217  equal.
MODEL RESULTS


                                      Two-Tailed –Factor loadings are now constrained to be equal.
                 Estimate       S.E.  Est./S.E.    P-Value
 SEPRE    BY
   SE1            1.000      0.000    999.000    999.000
   SE2             0.993      0.031     32.282      0.000
   SE3             0.980      0.032     30.422      0.000
   SE4            1.024      0.032     31.682      0.000
 SEPOST   BY
```

SE1P	1.000	0.000	999.000	999.000
SE2P	0.993	0.031	32.282	0.000
SE3P	0.980	0.032	30.422	0.000
SE4P	1.024	0.032	31.682	0.000
SEPOST WITH				
SEPRE	0.322	0.034	9.480	0.000
Means				
SEPRE	3.634	0.027	132.532	0.000
SEPOST	3.515	0.031	113.345	0.000
Intercepts				
SE1	0.000	0.000	999.000	999.000
SE2	-0.091	0.116	-0.783	0.434
SE3	-0.094	0.121	-0.770	0.441
SE4	0.237	0.122	1.949	0.051
SE1P	0.000	0.000	999.000	999.000
SE2P	-0.134	0.113	-1.186	0.235
SE3P	-0.145	0.118	-1.232	0.218
SE4P	0.192	0.118	1.623	0.105
Variances				
SEPRE	0.735	0.041	17.887	0.000
SEPOST	0.806	0.047	17.111	0.000
Residual Variances				
SE1	0.988	0.038	25.884	0.000
SE2	0.986	0.038	26.082	0.000
SE3	1.118	0.041	27.152	0.000
SE4	0.909	0.037	24.716	0.000
SE1P	0.878	0.039	22.302	0.000
SE2P	0.816	0.037	21.962	0.000
SE3P	0.991	0.042	23.327	0.000
SE4P	0.880	0.040	22.071	0.000

Scalar Model for Longitudinal Invariance: Selected Input

```
MODEL:
 SEPRE BY SE1
              SE2 (1)
              SE3 (2)
              SE4 (3);
 SEPOST BY SE1P
              SE2P (1)
              SE3P (2)
              SE4P (3);
 [SEPRE@0 SEPOST]; – Constrain the first factor mean to zero to be able to estimate all intercepts.
 [SE1 SE1P] (4); – Place parentheses next to intercepts to constrain them to be equal.
 [SE2 SE2P] (5);
 [SE3 SE3P] (6);
 [SE4 SE4P] (7);
OUTPUT: MODINDICES(ALL 0);
```

Scalar Model for Longitudinal Invariance: Selected Output

Chi-Square Test of Model Fit – **Chi-square and dfs are used to conduct chi-square difference test with**

Value 39.048 **metric model.**

Degrees of Freedom 25 – **We gained 3 df because we constrained the intercepts to be equal.**

P-Value 0.0364

Two-Tailed –**Factor loadings are still constrained to be equal.**

	Estimate	S.E.	Est./S.E.	P-Value
SEPRE BY				
SE1	1.000	0.000	999.000	999.000
SE2	0.997	0.031	32.478	0.000
SE3	0.985	0.032	30.628	0.000
SE4	1.028	0.032	31.897	0.000

SEPOST BY

SE1P	1.000	0.000	999.000	999.000
SE2P	0.997	0.031	32.478	0.000
SE3P	0.985	0.032	30.628	0.000
SE4P	1.028	0.032	31.897	0.000

SEPOST WITH

SEPRE	0.320	0.034	9.480	0.000

Means

SEPRE	0.000	0.000	999.000	999.000	–**Pre-test factor mean is now constrained to be zero.**
SEPOST	-0.153	0.029	-5.268	0.000	–**Post-test factor mean now represents change from pre to post-test. Additionally, the p-value indicates whether this change is significant.**

Intercepts –**Intercepts are now constrained to be equal.**

SE1	3.649	0.025	148.942	0.000
SE2	3.514	0.024	144.412	0.000
SE3	3.460	0.025	138.848	0.000
SE4	3.954	0.025	160.692	0.000
SE1P	3.649	0.025	148.942	0.000
SE2P	3.514	0.024	144.412	0.000
SE3P	3.460	0.025	138.848	0.000
SE4P	3.954	0.025	160.692	0.000

Variances

SEPRE	0.731	0.041	17.906	0.000
SEPOST	0.801	0.047	17.124	0.000

Residual Variances

SE1	0.990	0.038	25.964	0.000
SE2	0.986	0.038	26.084	0.000
SE3	1.117	0.041	27.137	0.000
SE4	0.909	0.037	24.712	0.000
SE1P	0.880	0.039	22.376	0.000

SE2P	0.816	0.037	21.960	0.000
SE3P	0.990	0.042	23.310	0.000
SE4P	0.879	0.040	22.064	0.000

Strict Model for Longitudinal Invariance: Selected Input

```
MODEL:
 SEPRE BY SE1
             SE2 (1)
             SE3 (2)
             SE4 (3);
 SEPOST BY SE1P
             SE2P (1)
             SE3P (2)
             SE4P (3);
 [SEPRE@0 SEPOST];
  [SE1 SE1P] (4);
  [SE2 SE2P] (5);
  [SE3 SE3P] (6);
  [SE4 SE4P] (7);
SEPRE SEPOST (8);
OUTPUT: MODINDICES(ALL 0);
```

– List factor variances and place parentheses next to them to constrain to be equal.

Strict Model (factor variances) for Longitudinal Invariance: Selected Output

Chi-Square Test of Model Fit – **Chi-square and dfs are used to conduct chi-square difference test with**
Value 41.590 – **scalar model.**
Degrees of Freedom 26 – **we gained one df because we constrained the two variances to**
P-Value 0.0270 **be equal.**

MODEL RESULTS

Two-Tailed – **Factor loadings still constrained to be equal.**

	Estimate	S.E.	Est./S.E.	P-Value
SEPRE BY				
SE1	1.000	0.000	999.000	999.000
SE2	0.997	0.031	32.450	0.000
SE3	0.985	0.032	30.604	0.000
SE4	1.028	0.032	31.884	0.000
SEPOST BY				
SE1P	1.000	0.000	999.000	999.000
SE2P	0.997	0.031	32.450	0.000
SE3P	0.985	0.032	30.604	0.000
SE4P	1.028	0.032	31.884	0.000
SEPOST WITH				
SEPRE	0.318	0.034	9.484	0.000
Means				
SEPRE	0.000	0.000	999.000	999.000
SEPOST	-0.152	0.029	-5.271	0.000
Intercepts – **intercepts still constrained to be equal.**				
SE1	3.649	0.025	147.318	0.000
SE2	3.513	0.025	142.798	0.000
SE3	3.460	0.025	137.434	0.000
SE4	3.953	0.025	158.909	0.000
SE1P	3.649	0.025	147.318	0.000
SE2P	3.513	0.025	142.798	0.000

SE3P	3.460	0.025	137.434	0.000
SE4P	3.953	0.025	158.909	0.000
Variances – **variances now constrained to be equal.**				
SEPRE	0.762	0.037	20.347	0.000
SEPOST	0.762	0.037	20.347	0.000
Residual Variances				
SE1	0.987	0.038	25.941	0.000
SE2	0.983	0.038	26.056	0.000
SE3	1.115	0.041	27.121	0.000
SE4	0.905	0.037	24.686	0.000
SE1P	0.883	0.039	22.438	0.000
SE2P	0.819	0.037	22.023	0.000
SE3P	0.993	0.043	23.366	0.000
SE4P	0.882	0.040	22.139	0.000

Strict Model for Longitudinal Invariance (Residual Error Variances): Selected Input

```
MODEL:
 SEPRE BY SE1
          SE2 (1)
          SE3 (2)
          SE4 (3);
 SEPOST BY SE1P
          SE2P (1)
          SE3P (2)
          SE4P (3);
 [SEPRE@0 SEPOST];
 [SE1 SE1P] (4);
 [SE2 SE2P] (5);
 [SE3 SE3P] (6);
 [SE4 SE4P] (7);
SEPRE SEPOST (8);
```

```
SE1 SE1P (9); – List factor variances and place parentheses next to them to constrain to be equal.
 SE2 SE2P (10);
 SE3 SE3P (11);
 SE4 SE4P (12);
OUTPUT: MODINDICES(ALL 0);
```

Strict Model (error variances) for Longitudinal Invariance: Selected Output

Chi-Square Test of Model Fit – **Values to conduct chi-square difference test with strict factor variance**

Value 62.843 **model.**

Degrees of Freedom 30 **we gained four df because we constrained all errors to be equal.**

P-Value 0.0004

MODEL RESULTS

Two-Tailed – **Factor loadings still constrained to be equal.**

	Estimate	S.E.	Est./S.E.	P-Value
SEPRE BY				
SE1	1.000	0.000	999.000	999.000
SE2	0.997	0.031	32.238	0.000
SE3	0.985	0.032	30.455	0.000
SE4	1.029	0.032	31.785	0.000
SEPOST BY				
SE1P	1.000	0.000	999.000	999.000
SE2P	0.997	0.031	32.238	0.000
SE3P	0.985	0.032	30.455	0.000
SE4P	1.029	0.032	31.785	0.000
SEPOST WITH				
SEPRE	0.317	0.034	9.459	0.000
Means				
SEPRE	0.000	0.000	999.000	999.000
SEPOST	-0.152	0.029	-5.270	0.000

Intercepts – **Intercepts still constrained to be equal.**

SE1	3.648	0.025	147.913	0.000
SE2	3.514	0.024	143.778	0.000
SE3	3.461	0.025	138.040	0.000
SE4	3.953	0.025	159.001	0.000
SE1P	3.648	0.025	147.913	0.000
SE2P	3.514	0.024	143.778	0.000
SE3P	3.461	0.025	138.040	0.000
SE4P	3.953	0.025	159.001	0.000

Variances – **Factor variances still constrained to be equal.**

SEPRE	0.760	0.038	20.270	0.000
SEPOST	0.760	0.038	20.270	0.000

Residual Variances – **Error variances still constrained to be equal.**

SE1	0.944	0.029	32.889	0.000
SE2	0.913	0.028	32.714	0.000
SE3	1.063	0.031	34.610	0.000
SE4	0.896	0.028	31.608	0.000
SE1P	0.944	0.029	32.889	0.000
SE2P	0.913	0.028	32.714	0.000
SE3P	1.063	0.031	34.610	0.000
SE4P	0.896	0.028	31.608	0.000

Configural Model for Multi-Group Invariance: Input

```
TITLE: CONFIGURAL MODEL SE MS and HS
DATA: file is MSHSsurvey.dat;
VARIABLE:
    NAMES ARE ID SL GENDER GRADE  SE1 SE2 SE3 SE4 SE1P SE2P SE3P SE4P;
     USEVARIABLES ARE SL SE1-SE4 SE1P-SE4P;
     GROUPING IS SL (0=MS 1=HS); – This is how you tell Mplus you have a grouping variable.
     MISSING=ALL (999);
MODEL: – Model commands for all groups
 SEPOST BY SE1P
           SE2P
           SE3P
           SE4P;
 [SEPOST@0]; – Void the Mplus default to constrain factor mean of reference group by constraining the
latent mean for both groups.

MODEL HS: – Model specific commands for high school (reference group)
SEPOST BY  SE2P – Must tell Mplus to estimate factor loadings to override the default to constrain them
           SE3P  equal across groups.
           SE4P;
 [SE1P-SE4P]; – Void the Mplus default to constrain intercepts to be equal across groups.
 OUTPUT: MODINDICES(ALL 0);
```

Configural Model for Multi-Group Invariance: Selected Output

SUMMARY OF ANALYSIS

Number of groups 2 **– Always check to make sure your number of groups is right.**

Chi-Square Test of Model Fit **– Chi-square and df used to conduct chi-square difference tests**

Value	15.632
Degrees of Freedom	4
P-Value	0.0036

Chi-Square Contributions From Each Group **– in multi-group analyses you get the chi-square contribution from each group. These add up to the total chi-square listed above.**

MS	13.582
HS	2.050

MODEL RESULTS **– Model estimates are reported by group.**

Two-Tailed **– Middle school student estimates.**

	Estimate	S.E.	Est./S.E.	P-Value
Group MS				
SEPOST BY				
SE1P	1.000	0.000	999.000	999.000
SE2P	1.002	0.044	22.941	0.000
SE3P	1.003	0.046	21.595	0.000
SE4P	1.023	0.047	21.851	0.000
Means				
SEPOST	0.000	0.000	999.000	999.000
Intercepts				
SE1P	3.514	0.031	112.194	0.000
SE2P	3.356	0.031	109.439	0.000
SE3P	3.299	0.032	102.278	0.000
SE4P	3.790	0.032	119.841	0.000
Variances				
SEPOST	0.796	0.056	14.114	0.000
Residual Variances				

SE1P	0.886	0.041	21.493	0.000
SE2P	0.815	0.039	20.827	0.000
SE3P	0.982	0.044	22.218	0.000
SE4P	0.880	0.042	21.057	0.000

Group HS – **High school estimates.**

SEPOST BY				
SE1P	1.000	0.000	999.000	999.000
SE2P	0.935	0.052	17.948	0.000
SE3P	1.044	0.055	19.086	0.000
SE4P	1.066	0.056	18.958	0.000
Means				
SEPOST	0.000	0.000	999.000	999.000
Intercepts				
SE1P	3.524	0.036	98.470	0.000
SE2P	3.432	0.036	96.101	0.000
SE3P	3.483	0.036	95.987	0.000
SE4P	3.781	0.038	99.098	0.000
Variances				
SEPOST	0.599	0.052	11.447	0.000
Residual Variances				
SE1P	0.498	0.033	15.043	0.000
SE2P	0.571	0.035	16.483	0.000
SE3P	0.476	0.033	14.245	0.000
SE4P	0.568	0.038	15.090	0.000

Metric Model for Multi-Group Invariance: Input
MODEL: – Model commands for all groups. SEPOST **BY** SE1P SE2P SE3P SE4P**;** **[**SEPOST**@0];** **MODEL** HS**: – Here we deleted the request to estimate unique factor loadings for HS students.** **[**SE1P-SE4P**];** **OUTPUT: MODINDICES(ALL 0);**

Metric Model for Multi-Group Invariance: Selected Output
Chi-Square Test of Model Fit **– Chi-square and dfs used to conduct chi-square difference test with configural Model.** Value 18.771 Degrees of Freedom 7 **– We gained 3 dfs by constraining the factor loadings to be equal.** P-Value 0 0.0089 Chi-Square Contributions From Each Group MS 14.889 HS 3.882 MODEL RESULTS Two-Tailed Estimate S.E. Est./S.E. P-Value Group MS **– MS results.** SEPOST BY SE1P 1.000 0.000 999.000 999.000 SE2P 0.976 0.033 29.181 0.000 **– MS factor loadings are constrained to be equal with HS.** SE3P 1.023 0.035 28.829 0.000 SE4P 1.041 0.036 28.845 0.000

```
Means
   SEPOST          0.000     0.000   999.000   999.000
 Intercepts
   SE1P            3.514     0.031   112.226     0.000
   SE2P            3.356     0.030   110.384     0.000
   SE3P            3.299     0.032   101.808     0.000
   SE4P            3.790     0.032   119.326     0.000
 Variances
   SEPOST          0.791     0.048    16.342     0.000
 Residual Variances
  SE1P             0.890     0.040    22.245     0.000
   SE2P            0.833     0.038    21.999     0.000
   SE3P            0.971     0.043    22.686     0.000
   SE4P            0.869     0.040    21.540     0.000
```

Group HS – **HS results.**

```
 SEPOST   BY
   SE1P            1.000     0.000   999.000   999.000
   SE2P            0.976     0.033    29.181     0.000
   SE3P            1.023     0.035    28.829     0.000
   SE4P            1.041     0.036    28.845     0.000
```

– **HS factor loadings are constrained to be equal with MS.**

```
 Means
   SEPOST          0.000     0.000   999.000   999.000
 Intercepts
   SE1P            3.524     0.036    98.423     0.000
   SE2P            3.432     0.036    94.539     0.000
   SE3P            3.483     0.036    96.689     0.000
   SE4P            3.781     0.038    99.894     0.000
 Variances
   SEPOST          0.601     0.043    13.850     0.000
 Residual Variances
   SE1P            0.497     0.032    15.603     0.000
```

SE2P	0.559	0.034	16.571	0.000
SE3P	0.484	0.032	15.145	0.000
SE4P	0.577	0.036	15.940	0.000

Scalar Model for Multi-Group Invariance: Input

```
MODEL: – Model commands for all groups
 SEPOST BY SE1P
           SE2P
           SE3P
           SE4P;
 [SEPOST@0];
MODEL HS: – Here we deleted the request to estimate unique intercepts for HS students.
OUTPUT: MODINDICES(ALL 0);
```

Scalar Model for Multi-Group Invariance: Selected Output

```
Chi-Square Test of Model Fit – Chi-square and dfs used to conduct chi-square difference test with metric
                               model.
     Value                  41.523 –Our chi-square is quite large, suggesting we don't have intercept invariance.
     Degrees of Freedom         11–We gained 4 dfs from constraining intercepts.
     P-Value                0.0000
Chi-Square Contributions From Each Group
     MS                     25.598
     HS                     15.925

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

Group MS – **MS results**

SEPOST BY					
SE1P	1.000	0.000	999.000	999.000	– **MS factor loadings are constrained to be equal with HS**
SE2P	0.978	0.034	29.167	0.000	
SE3P	1.025	0.036	28.766	0.000	
SE4P	1.040	0.036	28.816	0.000	
Means					
SEPOST	0.000	0.000	999.000	999.000	
Intercepts					
SE1P	3.514	0.024	149.172	0.000	– **MS intercepts are constrained to be equal with HS**
SE2P	3.387	0.023	145.152	0.000	
SE3P	3.385	0.024	139.782	0.000	
SE4P	3.784	0.024	155.285	0.000	
Variances					
SEPOST	0.791	0.048	16.328	0.000	
Residual Variances					
SE1P	0.890	0.040	22.238	0.000	
SE2P	0.832	0.038	21.959	0.000	
SE3P	0.977	0.043	22.644	0.000	
SE4P	0.871	0.040	21.560	0.000	

Group HS – **HS results**

SEPOST BY					
SE1P	1.000	0.000	999.000	999.000	– **HS factor loadings are constrained to be equal with MS**
SE2P	0.978	0.034	29.167	0.000	
SE3P	1.025	0.036	28.766	0.000	
SE4P	1.040	0.036	28.816	0.000	
Means					
SEPOST	0.000	0.000	999.000	999.000	
Intercepts					
SE1P	3.514	0.024	149.172	0.000	–**HS intercepts are constrained to be equal with MS**

```
   SE2P             3.387      0.023    145.152      0.000
   SE3P             3.385      0.024    139.782      0.000
   SE4P             3.784      0.024    155.285      0.000
 Variances
   SEPOST             0.602      0.043      13.838      0.000
 Residual Variances
   SE1P             0.497      0.032     15.581      0.000
   SE2P             0.558      0.034     16.533      0.000
   SE3P             0.490      0.032     15.112      0.000
   SE4P             0.579      0.036     15.938      0.000
```

MODEL MODIFICATION INDICES – **Modification indices tell us when parameter constraints are poorly chosen.**

```
Minimum M.I. value for printing the modification index      0.000
                        M.I.     E.P.C.  Std E.P.C.  StdYX E.P.C.
Group MS
Means/Intercepts/Thresholds
[ SE1P     ]             2.027     0.024      0.024       0.019
[ SE2P     ]             0.664    -0.013     -0.013      -0.010
[ SE3P     ]            18.733    -0.079     -0.079      -0.059
[ SE4P     ]             4.147     0.033      0.033       0.025
```

–If we release the intercept of item 3 our chi-square will go down by appox 19.

```
Group MS
[ SE1P     ]             2.027    -0.029     -0.029      -0.028
[ SE2P     ]             0.664     0.018      0.018       0.017
[ SE3P     ]            18.731     0.086      0.086       0.081
[ SE4P     ]             4.146    -0.045     -0.045      -0.041
```

–If we release the intercept of item 3 our chi-square will go down by appox 19.

Strict Model (factor variances) for Multi-Group Invariance: Input
***There was a significant difference in model fit between the metric and scalar models. In the PowerPoint we demonstrated that we are able to attain partial invariance of intercepts. However, for this next example we are just going to assume we attained full intercept invariance.**
MODEL: – **Model commands for all groups**
SEPOST **BY** SE1P
SE2P
SE3P
SE4P**;**
[SEPOST**@0];**
SEPOST (1); – **Here we are constraining the factor variances to be equal**
MODEL HS**:**
OUTPUT: MODINDICES(ALL 0);

Strict Model (factor variances) for Multi-Group Invariance: Selected Output
Chi-Square Test of Model Fit –**Chi square and df to conduct chi square difference test with scalar model.**
Value 55.139 –**Here we can see our chi-square is high suggesting non invariance.**
Degrees of Freedom 12 –**we gained one df from constraining the factor variances.**
P-Value 0.0000
Chi-Square Contributions From Each Group
MS 30.272
HS 24.866
MODEL RESULTS
Two-Tailed
Estimate S.E. Est./S.E. P-Value
Group MS –**MS estimates**
SEPOST BY –**factor loadings constrained to be equal with HS factor loadings.**

```
  SE1P        1.000    0.000  999.000  999.000
  SE2P        0.979    0.034   29.130    0.000
  SE3P        1.025    0.036   28.852    0.000
  SE4P        1.043    0.036   28.851    0.000
 Means
  SEPOST      0.000    0.000  999.000  999.000
```

Intercepts –**Intercepts are constrained to be equal HS intercepts**

```
  SE1P        3.511    0.024  148.634    0.000
  SE2P        3.384    0.023  144.639    0.000
  SE3P        3.382    0.024  139.251    0.000
  SE4P        3.781    0.024  154.721    0.000
```

Variances–**Factor variance now constrained to be equal with HS factor variance.**

```
  SEPOST      0.717    0.040   18.089    0.000
 Residual Variances
  SE1P        0.897    0.040   22.343    0.000
  SE2P        0.838    0.038   22.052    0.000
  SE3P        0.982    0.043   22.729    0.000
  SE4P        0.875    0.041   21.607    0.000
```

Group HS–**HS estimates**

SEPOST BY –**factor loadings constrained to be equal with MS factor loadings.**

```
  SE1P        1.000    0.000  999.000  999.000
  SE2P        0.979    0.034   29.130    0.000
  SE3P        1.025    0.036   28.852    0.000
  SE4P        1.043    0.036   28.851    0.000
 Means
 SEPOST       0.000    0.000  999.000  999.000
```

Intercepts –**Intercepts constrained to be equal with MS intercepts.**

```
  SE1P        3.511    0.024  148.634    0.000
  SE2P        3.384    0.023  144.639    0.000
  SE3P        3.382    0.024  139.251    0.000
```

SE4P	3.781	0.024	154.721	0.000
Variances **–Factor variance now constrained to be equal with MS factor variance.**				
SEPOST	0.717	0.040	18.089	0.000
Residual Variances				
SE1P	0.494	0.032	15.518	0.000
SE2P	0.557	0.034	16.483	0.000
SE3P	0.486	0.032	15.042	0.000
SE4P	0.574	0.036	15.848	0.000

Strict Model (residual error variances) for Multi-Group Invariance: Input

***We did not attain invariance of factor variances. However, for this next example we are just going to assume we attained invariance of factor variances so we can assess residual error invariance.**

MODEL: – **Model commands for all groups**

```
 SEPOST BY SE1P
            SE2P
            SE3P
            SE4P;
 [SEPOST@0];
 SEPOST (1); –Residual error variances listed and constrained to be equal.
 SE1P (2);
 SE2P (3);
 SE3P (4);
 SE4P (5);
MODEL HS:
OUTPUT: MODINDICES(ALL 0);
```

Strict Model (residual error variances) for Multi-Group Invariance: Selected Output

Chi-Square Test of Model Fit –**Chi square and df to conduct chi-square difference test with factor variance model**

Value 291.701–**Here we can see our chi-square is huge suggesting invariance.**

Degrees of Freedom 16–**we gained 4 dfs from constraining the error variances.**

P-Value 0.0000

Chi-Square Contributions From Each Group

MS 84.667

HS 207.034

Group MS–**MS estimates**

SEPOST BY–**MS factor loadings still constrained to be equal**

SE1P	1.000	0.000	999.000	999.000
SE2P	0.984	0.034	28.826	0.000
SE3P	1.017	0.036	27.947	0.000
SE4P	1.034	0.037	28.158	0.000

Means

SEPOST	0.000	0.000	999.000	999.000

Intercepts –**MS intercepts still constrained to be equal**

SE1P	3.517	0.024	146.244	0.000
SE2P	3.381	0.024	142.879	0.000
SE3P	3.360	0.025	135.911	0.000
SE4P	3.787	0.025	153.833	0.000

Variances –**MS factor variance still constrained to be equal**

SEPOST	0.730	0.041	17.896	0.000

Residual Variances – **MS Residual variances now constrained to be equal**

SE1P	0.758	0.029	26.249	0.000
SE2P	0.736	0.028	26.365	0.000
SE3P	0.817	0.031	26.677	0.000

```
  SE4P        0.777    0.030    25.863    0.000
Group HS–HS estimates
 SEPOST   BY–HS factor loadings still constrained to be equal
  SE1P        1.000    0.000   999.000  999.000
  SE2P        0.984    0.034    28.826    0.000
  SE3P        1.017    0.036    27.947    0.000
  SE4P        1.034    0.037    28.158    0.000
 Means
  SEPOST        0.000    0.000   999.000  999.000
 Intercepts–HS intercepts still constrained to be equal
  SE1P        3.517    0.024   146.244    0.000
  SE2P        3.381    0.024   142.879    0.000
  SE3P        3.360    0.025   135.911    0.000
  SE4P        3.787    0.025   153.833    0.000
 Variances –HS factor variances still constrained to be equal
  SEPOST        0.730    0.041    17.896    0.000
 Residual Variances–HS residual variances now constrained to be equal
  SE1P        0.758    0.029    26.249    0.000
  SE2P        0.736    0.028    26.365    0.000
  SE3P        0.817    0.031    26.677    0.000
  SE4P        0.777    0.030    25.863    0.000
```

Section 3: *Software Options*

Lavann

Website: http://lavaan.ugent.be/
Online Support: https://groups.google.com/forum/#!forum/lavaan
Pricing: Free
Notes: Highly recommended on online SEM forums

R

Website: http://www.r-project.org/
Online Support: http://blog.revolutionanalytics.com/local-r-groups.html
Pricing: Free

EQS

Website: http://www.mvsoft.com/
Pricing: Free
Online support resources: http://www.mvsoft.com/techsup.htm

Mplus

Website: http://www.statmodel.com/
Online Support: http://www.statmodel.com/cgi-bin/discus/discus.cgi
Pricing: Lifetime membership with upgrades included (range represents different packages available). Student: $195-$350, University Pricing: $595-895, Commercial/non-profit/govt: $695-1,095
Notes: Highly recommend by workshop facilitators, technical support from creators of software (usually within 24 hours).

Amos

Website: http://www-03.ibm.com/software/products/us/en/spss-amos/
Pricing: Approximately $1,590 (exact pricing range unknown)
Notes: Rated highly for its graphical component

LISREL

Website: http://www.ssicentral.com/lisrel/
Online Support: http://www.ssicentral.com/lisrel/resources.html
Pricing: Single User $495, 12 month rental $130

Helpful discussion thread on strengths and weaknesses of different software options:
http://www.researchgate.net/post/What_is_your_favorite_Structural_Equation_Modeling_program

Section 4: *Additional Resources*

Primary text recommended by facilitators:

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications and programming.* New York, NY: Taylor & Francis.

Other Resources:

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Byrne, B. M. (1994). Testing for the factorial validity, replication, and invariance of a measuring instrument: A paradigmatic application based on the Maslach Burnout Inventory. Multivariate Behavioral Research, 29, 289-311.

Byrne B. M. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology, 34*(2), 155-175.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups?: Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care, 44*(11 Suppl 3), s78-s94.

Hu, L-T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). Thousand Oaks, CA: Sage.

Hu, L-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1,* 130-149.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Pyschometrika, 58*, 525-543.

Meredith, W., & Teresi, J. A. (2013). An essay on measurement and factorial invariance. *Medical Care, 44*(11 Suppl 3), s69-s77.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In J.A. Bollen & J. S. Long (Eds.), *testing structural equation models* (pp.10-39). Newbury park: CA: Sage.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, recommendations, for organizational research. *Organizational Research Methods, 3*(1), 4-70.

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, May, 1-7.*

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.) *The science of prevention: Methodological advances from alcohol and substance abuse research*, pp. 281-324. Washington, DC: APA.