Electronic Supplementary Information

Addressing diversity and inclusion through group comparisons: A primer on measurement invariance testing.

Guizella A. Rocabado,[1] Regis Komperda,[2‡] Jennifer E. Lewis[1,3] and Jack Barbera[4†]

The purpose of the electronic supplementary information (ESI) is to provide readers with the data and code necessary to reproduce the examples from the main body of the paper as well as to provide a template for conducting invariance testing on a simulated data set that can be modified for those interested in conducting invariance testing on their own data. The code in the ESI is primarily written for the R statistical computing language, though Mplus code is also included for conducting invariance testing. The code in the ESI is also available through GitHub (https://github.com/RegisBK/Invariance_CERP) as this provides an easier way to download and use the code rather than cutting and pasting from this document. All analyses were conducted with R version 3.6.1 (R Core Team, 2019) and Mplus version 8.2.

This document assumes a basic understanding of how to work with R and/or Mplus. Users less familiar with these programs are encouraged to consult any of the resources available describing the use of these programs (Hirschfeld and Von Brachel, 2014; Komperda, 2017; Muthén and Muthén, 2017; Rosseel, 2020). Unless otherwise noted, the code provided here is intended to be entered directly into the software and is written in a different font to distinguish it from explanatory text.

Table of Contents

---

[1] Department of Chemistry, University of South Florida, USA.
[2] Department of Chemistry & Biochemistry; Center for Research in Mathematics and Science Education, San Diego State University, USA.
[3] Center for the Improvement of Teaching and Research in Undergraduate STEM Education, University of South Florida, USA
[4] Department of Chemistry, Portland State University, USA.
† Corresponding author for manuscript (jbarbera@pdx.edu)
‡ Corresponding author for ESI (rkomperda@sdsu.edu)
See DOI: 10.1039/d0rp00025f

Simulation and Visualization of Data in R

Simulation of Identical Group Data

The data used for the examples in the main article are simulated data created in R to follow the structure of the fictional Perceived Relevance of Chemistry Questionnaire (PRCQ). The PRCQ is conceptualized as containing three fictitious subconstructs: Importance of Chemistry (IC), Connectedness of Chemistry (CC), and Applications of Chemistry (AC). Additionally, the fictitious PRCQ is designed to be a 12-item instrument, where there are four items designed to measure each of the three subconstructs. To simulate this data in R first requires the installation and loading of the package `simstandard` (Schneider, 2019) which requires other dependent packages such as `dplyr` (Wickham *et al.*, 2019) to be installed as well.

```
install.packages("simstandard")
library(simstandard)
```

Syntax from the `lavaan` factor analysis package (Rosseel, 2012) is used to specify a three-factor model with four items associated with each factor. For this model, named `PRCQ`, items 1–4 are associated with the IC factor, 5–8 with the CC factor, and 9–12 with the AC factor. All items are assigned to have the same strength of association with their respective factors, a standardized value of 0.8. This value was chosen as it is relatively strong but not perfect association. In addition, each factor was simulated as having a weak association with the other factors. IC and CC have an association of 0.3, IC and AC have an association of 0.2 and CC and AC have an association of 0.1.

```
PRCQ<-'
      IC =~ 0.8*I1 + 0.8*I2 + 0.8*I3 + 0.8*I4
      CC =~ 0.8*I5 + 0.8*I6 + 0.8*I7 + 0.8*I8
      AC =~ 0.8*I9 + 0.8*I10 + 0.8*I11 + 0.8*I12

      IC  =~  0.3*CC
      IC  =~  0.2*AC
      CC  =~  0.1*AC
'
```

Now, observed data that follow the relations described by the model can be simulated. The `set.seed()` function is used to ensure reproducibility across uses by simulating the same pseudorandom data each time the code is run. Following the example from the main text, data are simulated separately for 1000 fictional students in the STEM majors group and for 1000 students in the non-STEM majors group. A column named `group` is added to distinguish the data from each group and the two datasets are combined to form the new dataset named `combined`.

```
set.seed(1234)
STEM <- sim_standardized(PRCQ, n = 1000, observed = T, latent = F,
errors = F)
nonSTEM <- sim_standardized(PRCQ, n = 1000, observed = T, latent = F,
errors = F)
```

```
STEM$group<-"STEM"
nonSTEM$group<-"nonSTEM"

combined<-rbind(STEM, nonSTEM)
```

The data generated with `sim_standardized()` are standardized meaning they have an average value of 0 and standard deviation of 1 as well as a normal distribution. Descriptive statistics for the complete dataset and for each group within the dataset can be generated using the `describe()` and `describeBy()` functions in the `psych` package (Revelle, 2018) and are shown in Figure ESI1 and ESI2. Note that statistics are not generated for the `group` variable as it is a character, not a number.

```
library(psych)
describe(combined)
describeBy(combined, group="group")
```

```
> describe(combined)
       vars    n  mean   sd median trimmed  mad   min  max range  skew kurtosis   se
I1        1 2000 -0.01 0.98  -0.01   -0.01 0.97 -3.26 3.22  6.48  0.03    -0.02 0.02
I2        2 2000  0.00 1.00  -0.06    0.00 1.01 -3.44 3.44  6.88  0.02    -0.05 0.02
I3        3 2000 -0.03 0.99  -0.02   -0.03 0.98 -3.22 3.07  6.28  0.01    -0.05 0.02
I4        4 2000 -0.03 1.01   0.00   -0.03 1.02 -3.06 3.47  6.53  0.03    -0.12 0.02
I5        5 2000 -0.02 0.98  -0.01   -0.01 0.97 -3.13 3.39  6.52 -0.08    -0.06 0.02
I6        6 2000 -0.02 0.99  -0.01   -0.02 1.00 -3.53 3.36  6.89 -0.03     0.04 0.02
I7        7 2000  0.00 0.99   0.01    0.00 1.01 -3.03 4.22  7.24 -0.01     0.04 0.02
I8        8 2000  0.00 1.00   0.00    0.01 1.01 -3.63 3.03  6.66 -0.06    -0.03 0.02
I9        9 2000  0.01 0.97   0.03    0.01 0.97 -3.05 3.61  6.65  0.01     0.02 0.02
I10      10 2000  0.03 1.00   0.02    0.02 1.00 -2.95 3.40  6.35  0.11    -0.07 0.02
I11      11 2000  0.02 0.98   0.05    0.01 0.95 -2.88 3.48  6.36  0.04     0.00 0.02
I12      12 2000  0.01 0.98   0.01    0.00 0.98 -3.58 4.11  7.69  0.15     0.11 0.02
group*   13 2000   NaN   NA     NA     NaN   NA   Inf -Inf  -Inf    NA       NA   NA
```

Figure ESI1. Output from the `describe()` function using the dataset named `combined`.

```
> describeBy(combined, group="group")

 Descriptive statistics by group
group: nonSTEM
       vars    n  mean   sd median trimmed  mad   min  max range  skew kurtosis   se
I1        1 1000 -0.01 1.00  -0.04   -0.01 0.99 -3.26 3.15  6.42  0.06    -0.06 0.03
I2        2 1000 -0.02 1.01  -0.08   -0.02 1.00 -3.44 2.84  6.28 -0.03     0.02 0.03
I3        3 1000 -0.03 0.98  -0.01   -0.03 0.99 -2.90 3.07  5.97  0.01    -0.06 0.03
I4        4 1000 -0.05 1.02  -0.03   -0.05 1.06 -3.06 3.47  6.53  0.05    -0.17 0.03
I5        5 1000 -0.02 0.98  -0.01   -0.01 0.97 -3.13 3.39  6.52 -0.13     0.09 0.03
I6        6 1000 -0.05 1.01  -0.02   -0.04 1.03 -3.53 3.19  6.73 -0.12    -0.06 0.03
I7        7 1000 -0.04 1.01  -0.02   -0.03 1.04 -3.03 3.15  6.17 -0.05    -0.07 0.03
I8        8 1000 -0.01 1.02   0.00    0.00 1.00 -3.63 2.98  6.61 -0.08     0.12 0.03
I9        9 1000  0.04 0.97   0.06    0.04 0.97 -3.05 3.61  6.65  0.05     0.13 0.03
I10      10 1000  0.07 0.99   0.05    0.06 1.01 -2.74 3.08  5.82  0.12    -0.16 0.03
I11      11 1000  0.04 0.98   0.06    0.03 1.00 -2.66 3.35  6.00  0.10    -0.17 0.03
I12      12 1000  0.05 0.97   0.04    0.04 0.98 -3.58 4.11  7.69  0.15     0.35 0.03
group*   13 1000   NaN   NA     NA     NaN   NA   Inf -Inf  -Inf    NA       NA   NA
-------------------------------------------------------------------------------------
group: STEM
       vars    n  mean   sd median trimmed  mad   min  max range  skew kurtosis   se
I1        1 1000  0.00 0.97   0.01    0.00 0.92 -3.02 3.22  6.23  0.00     0.01 0.03
I2        2 1000  0.01 0.99  -0.05    0.01 1.04 -2.74 3.44  6.18  0.07    -0.15 0.03
I3        3 1000 -0.03 1.00  -0.04   -0.03 0.99 -3.22 3.06  6.28  0.02    -0.06 0.03
I4        4 1000  0.00 1.00   0.00   -0.01 0.98 -3.05 3.21  6.26  0.02    -0.07 0.03
I5        5 1000 -0.01 0.98  -0.02   -0.01 0.99 -2.95 2.95  5.90 -0.03    -0.21 0.03
I6        6 1000  0.00 0.98  -0.01    0.00 0.96 -3.22 3.36  6.58  0.07     0.12 0.03
I7        7 1000  0.03 0.97   0.03    0.03 0.97 -2.91 4.22  7.12  0.05     0.13 0.03
I8        8 1000  0.01 0.98   0.02    0.02 1.02 -3.03 3.03  6.06 -0.03    -0.23 0.03
I9        9 1000 -0.02 0.97   0.00   -0.01 0.98 -2.79 2.89  5.68 -0.03    -0.11 0.03
I10      10 1000 -0.01 1.00  -0.03   -0.01 0.99 -2.95 3.40  6.35  0.10     0.00 0.03
I11      11 1000  0.00 0.97   0.04    0.00 0.90 -2.88 3.48  6.36 -0.02     0.15 0.03
I12      12 1000 -0.02 0.99  -0.03   -0.03 0.98 -2.67 3.14  5.81  0.15    -0.12 0.03
group*   13 1000   NaN   NA     NA     NaN   NA   Inf -Inf  -Inf    NA       NA   NA
```

Figure ESI2. Output by group from the `describeBy()` function using the dataset named `combined`.

Additionally, the data are complete with no missing cases. These data may not be representative of the type of data obtained in chemistry education research using a non-fictional assessment instrument. For the purposes of this example, as in the main body of the text, this dataset will continue to be used. Further procedures in the ESI will demonstrate converting the data from continuous into categorical, which may better match authentic data.

Simulation of Data with Unequal Factor Loadings and Unequal Item Means

The previous section described the simulation of data for two groups using the same model in each group. To illustrate the effect of invariance at different levels, modifications were made to the data. The data are simulated to highlight specific issues that could be encountered (i.e., noninvariant loadings, noninvariant intercepts) but are unlikely to be representative of authentic data which could have numerous issues simultaneously. The model below is used to simulate data with a lower association between AC and I10 for the non-STEM majors group (changed to 0.3 instead of 0.8), as used to generate Figure 4 in the manuscript. This data is combined with the original STEM majors data to create the `combined.invar.load` dataset.

```
PRCQ.invar.load<-'
      IC =~ 0.8*I1 + 0.8*I2 + 0.8*I3 + 0.8*I4
      CC =~ 0.8*I5 + 0.8*I6 + 0.8*I7 + 0.8*I8
      AC =~ 0.8*I9 + 0.3*I10 + 0.8*I11 + 0.8*I12

      IC  =~  0.3*CC
      IC  =~  0.2*AC
      CC  =~  0.1*AC
'

nonSTEM.invar.load <- sim_standardized(PRCQ.invar.load, n = 1000,
observed = T, latent = F, errors = F)

nonSTEM.invar.load$group<-"nonSTEM"

combined.invar.load<-rbind(STEM, nonSTEM.invar.load)
```

To create data with a higher mean for I3 in the STEM majors group, as used to generate Figures 4 and 5 in the manuscript, a new dataset is created from the original STEM majors data and constant of 2 is added to all values for I3 in this new data. The STEM majors data is combined with the original non-STEM majors data to create a `combined.invar.mean` dataset. The `describeBy()` function can be used to confirm differences between the groups as seen in the descriptive statistics in Figure ESI3.

```
STEM.invar.mean<-STEM
STEM.invar.mean$I3<-STEM.invar.mean$I3+2

STEM.invar.mean$group<-"STEM"

combined.invar.mean<-rbind(STEM.invar.mean, nonSTEM)

describeBy(combined.invar.mean, group="group")
```

```
> describeBy(combined.invar.mean, group="group")

 Descriptive statistics by group
group: nonSTEM
       vars    n  mean   sd median trimmed  mad   min  max range  skew kurtosis   se
I1        1 1000 -0.01 1.00  -0.04   -0.01 0.99 -3.26 3.15  6.42  0.06    -0.06 0.03
I2        2 1000 -0.02 1.01  -0.08   -0.02 1.00 -3.44 2.84  6.28 -0.03     0.02 0.03
I3        3 1000 -0.03 0.98  -0.01   -0.03 0.99 -2.90 3.07  5.97  0.01    -0.06 0.03
I4        4 1000 -0.05 1.02  -0.03   -0.05 1.06 -3.06 3.47  6.53  0.05    -0.17 0.03
I5        5 1000 -0.02 0.98  -0.01   -0.01 0.97 -3.13 3.39  6.52 -0.13     0.09 0.03
I6        6 1000 -0.05 1.01  -0.02   -0.04 1.03 -3.53 3.19  6.73 -0.12    -0.06 0.03
I7        7 1000 -0.04 1.01  -0.02   -0.03 1.04 -3.03 3.15  6.17 -0.05    -0.07 0.03
I8        8 1000 -0.01 1.02   0.00    0.00 1.00 -3.63 2.98  6.61 -0.08     0.12 0.03
I9        9 1000  0.04 0.97   0.06    0.04 0.97 -3.05 3.61  6.65  0.05     0.13 0.03
I10      10 1000  0.07 0.99   0.05    0.06 1.01 -2.74 3.08  5.82  0.12    -0.16 0.03
I11      11 1000  0.04 0.98   0.06    0.03 1.00 -2.66 3.35  6.00  0.10    -0.17 0.03
I12      12 1000  0.05 0.97   0.04    0.04 0.98 -3.58 4.11  7.69  0.15     0.35 0.03
group*   13 1000   NaN   NA     NA     NaN   NA   Inf -Inf  -Inf    NA       NA   NA
------------------------------------------------------------------------------------
group: STEM
       vars    n  mean   sd median trimmed  mad   min  max range  skew kurtosis   se
I1        1 1000  0.00 0.97   0.01    0.00 0.92 -3.02 3.22  6.23  0.00     0.01 0.03
I2        2 1000  0.01 0.99  -0.05    0.01 1.04 -2.74 3.44  6.18  0.07    -0.15 0.03
I3        3 1000  1.97 1.00   1.96    1.97 0.99 -1.22 5.06  6.28  0.02    -0.06 0.03
I4        4 1000  0.00 1.00   0.00   -0.01 0.98 -3.05 3.21  6.26  0.02    -0.07 0.03
I5        5 1000 -0.01 0.98  -0.02   -0.01 0.99 -2.95 2.95  5.90 -0.03    -0.21 0.03
I6        6 1000  0.00 0.98  -0.01    0.00 0.96 -3.22 3.36  6.58  0.07     0.12 0.03
I7        7 1000  0.03 0.97   0.03    0.03 0.97 -2.91 4.22  7.12  0.05     0.13 0.03
I8        8 1000  0.01 0.98   0.02    0.02 1.02 -3.03 3.03  6.06 -0.03    -0.23 0.03
I9        9 1000 -0.02 0.97   0.00   -0.01 0.98 -2.79 2.89  5.68 -0.03    -0.11 0.03
I10      10 1000 -0.01 1.00  -0.03   -0.01 0.99 -2.95 3.40  6.35  0.10     0.00 0.03
I11      11 1000  0.00 0.97   0.04    0.00 0.90 -2.88 3.48  6.36 -0.02     0.15 0.03
I12      12 1000 -0.02 0.99  -0.03   -0.03 0.98 -2.67 3.14  5.81  0.15    -0.12 0.03
group*   13 1000   NaN   NA     NA     NaN   NA   Inf -Inf  -Inf    NA       NA   NA
```

Figure ESI3. Output by group from the `describeBy()` function using the dataset named `combined.invar.mean` showing different means for I3 across groups.

Visualization of Data

The R code in this section can be used to generate the data visualizations (correlations and distributions) shown in Figures 1–5 of the manuscript. Correlation plots can be made with the `corrplot` package (Wei and Simko, 2017). To use the `corrplot()` function, the numeric variables are selected from the `combined` dataset and a correlation matrix is generated with the `cor()` function. Additional function arguments are used to specify that colored boxes should be plotted (`method="color"`), the text should be in the diagonal of the matrix in black (`tl.pos="d", tl.col="black"`), only the lower diagonal of the correlation matrix should be visualized (`type="lower"`), and that grey grid lines should appear (`addgrid.col="grey"`). Specifying the size of the margins is done to make room for the plot title (`mar=c(0,0,1,0)`).

```
library(dplyr)
library(corrplot)

combined %>% select(I1:I12) %>%   cor() %>%
   corrplot(., method="color", tl.pos="d", tl.col="black",
   type="lower", addgrid.col="grey", mar=c(0,0,1,0))
```

Similar plots can be generated for subsets of the data by filtering the `combined` dataset using the group variable (`filter(group=="STEM")`).

```
combined %>% filter(group=="STEM") %>% select(I1:I12) %>% cor() %>%
   corrplot(., method="color", tl.pos="d", tl.col="black",
   type="lower", addgrid.col="grey", title="STEM Majors",
   mar=c(0,0,1,0))

combined %>% filter(group=="nonSTEM") %>% select(I1:I12) %>% cor() %>%
   corrplot(., method="color", tl.pos="d", tl.col="black",
   type="lower", addgrid.col="grey",title="Non-STEM Majors",
   mar=c(0,0,1,0))
```

Using the `combined.invar.load` dataset will produce Figure 3 images from the manuscript.

```
combined.invar.load %>% select(I1:I12) %>%   cor() %>%
    corrplot(., method="color", tl.pos="d", tl.col="black",
    type="lower", addgrid.col="grey",
    title="Combined Data Varied\n Strength of Association for I10",
   mar=c(0,0,1,0))

combined.invar.load %>% filter(group=="STEM") %>% select(I1:I12) %>%
   cor() %>% corrplot(., method="color", tl.pos="d", tl.col="black",
   type="lower", addgrid.col="grey", title="STEM Majors",
   mar=c(0,0,1,0))

combined.invar.load %>% filter(group=="nonSTEM")%>% select(I1:I12) %>%
   cor() %>% corrplot(., method="color", tl.pos="d", tl.col="black",
   type="lower", addgrid.col="grey",title="Non-STEM Majors",
   mar=c(0,0,1,0))
```

The Figure 4 images from the manuscript are produced using the same method with the `combined.invar.mean` dataset.

```
combined.invar.mean %>% select(I1:I12) %>%   cor() %>%
    corrplot(., method="color", tl.pos="d", tl.col="black",
   type="lower", addgrid.col = "grey",
   title="Combined Data\n Varied Mean for I3",mar=c(0,0,1,0))

combined.invar.mean %>% filter(group=="STEM") %>% select(I1:I12) %>%
  cor() %>% corrplot(., method="color", tl.pos="d", tl.col="black",
  type="lower", addgrid.col = "grey", title="STEM Majors",
  mar=c(0,0,1,0))
```

```
combined.invar.mean %>% filter(group=="nonSTEM") %>% select(I1:I12)
   %>%  cor() %>% corrplot(., method="color", tl.pos="d",
   tl.col="black", type="lower", addgrid.col = "grey",
   title="Non-STEM Majors", mar=c(0,0,1,0))
```

In order to generate the boxplot Figure 5 of the manuscript the package `reshape2` (Wickham, 2007) is needed to restructure the dataset and the package `ggplot2` (Wickham, 2016) is used to create the plot. First, the STEM and non-STEM groups are given more descriptive names since those will appear in the figure legend. The groups are also ordered as with STEM Majors first since the default setting would put the groups in alphabetical order.

```
library(ggplot2)
library(reshape2)

combined.invar.mean$group<-ifelse(combined.invar.mean$group=="STEM",
                                   "STEM Majors", "Non-STEM Majors")
combined.invar.mean$group<-ordered(combined.invar.mean$group,
                              levels=c("STEM Majors", "Non-STEM Majors"))
```

Next, the `melt()` function is used to create a long-format dataset where each group, variable (Item), and value occupies a single column. This long format is necessary for plotting using the function `ggplot()` with `geom_boxplot()`. In this boxplot the x-axis is the group and the y-axis is the value for each variable (`x=group, y=value, fill=group`). Faceting by variable (`facet_grid(.~variable)`) plots each item separately, yet within a single plot. The remainder of the code provides graphical parameters.

```
melt.mean<-combined.invar.mean %>%
                        select(I1:I12, group) %>% melt(id="group")
melt.mean$group<-melt.mean$group %>% as.factor()

ggplot(melt.mean, aes(x=group, y=value, fill=group))+
  geom_boxplot() + facet_grid(.~variable) + theme_bw() +
  theme(axis.title.x=element_blank(), axis.text.x=element_blank(),
        axis.ticks.x=element_blank(), axis.title.y=element_blank(),
        legend.position="bottom") +
  scale_fill_discrete(name="Group")
```

Conducting Invariance Testing

This section provides an overview of how to conduct measurement invariance testing using two popular software platforms, R and Mplus. Results obtained from both pieces of software will be similar, so the selection of software depends on the preferences of the researcher. In addition to R and Mplus there are other tools available for conducting measurement invariance testing, including SAS, LISREL, EQS, or the AMOS add-in for SPSS. A helpful comparison of software for structural equation modeling with multiple groups can be found in Narayana (2012) and Byrne (2004) provides a guide to AMOS.

Before introducing the specific steps to take within R and Mplus, it is worthwhile to note the default settings of both software packages. Within R, the package `lavaan` is generally used for factor analyses and in this package the default way to provide scale to the factor is to fix the value of the first item loading to one. In Mplus, the factor is given scale by setting its variance to one. Both methods are acceptable ways of identifying the model and will give equivalent results. However, each of these methods has different implications in the context of measurement invariance testing with multiple groups.

The method of setting the factor variance to one (as in Mplus) in both groups is generally not recommended for multigroup measurement invariance testing as it implies that the latent variable has the same variance in both groups. This is described as homogeneity of variance for the latent variables. Though conceptually similar to the test for homogeneity of variance used in *t*-tests and ANOVAs, in a latent framework this is an untestable assumption (Hancock *et al.*, 2009, 168).

In the first method, used within `lavaan,` setting an item loading to one, the default is to use the first item on the scale. When the first item on the scale is set to be one for both groups the rest of the series of structural equations will be solved assuming this item has the same loading value in both groups. Yet, there is no way to know for certain if that assumption is true or if there are other scale items that would have been better to set equivalent. This seemingly inconsequential decision can have major implications for interpretation of results and researchers are advised to think carefully about which item may be best to set equal across groups based on either theoretical or observable grounds (Bontempo and Hofer, 2007; Hancock *et al.*, 2009).

Invariance Testing with R – Continuous Data

Within the R software, the package `lavaan`, previously used to generate the simulated data, can be used to test confirmatory factor (CFA) models as well as structural equation models (SEM). The function for performing CFA, `cfa()` contains built-in arguments to set various model parameters equal for invariance testing (Hirschfeld and Von Brachel, 2014), making invariance testing a relatively simple process. In this section, the steps for measurement invariance testing will follow those in the main article using the `combined.invar.load` dataset to generate the fit index data from Table 1 in the manuscript. The general process for invariance testing within R is that of building up from the least constrained model (i.e., configural invariance) to the most constrained model (i.e., conservative invariance). Identical steps can be followed for the other datasets and fit indices resulting from these tests are provided later sections.

Step 0: Establishing Baseline Model

Following the steps outlined in the manuscript, the baseline model is tested for each group separately. The model is specified in the same manner as was used to generate the simulated data with the main difference being that values for the loadings and associations between factors are not assigned but will be estimated by the software from the data. This model is named `model.test` to distinguish it from the model used to simulate the data.

```
library(lavaan)

model.test<-'
      IC =~ I1 + I2 + I3 + I4
      CC =~ I5 + I6 + I7 + I8
      AC =~ I9 + I10 + I11 + I12
      '
```

The function `cfa()` is now used to examine how well the data fit the proposed model. The maximum likelihood (ML) estimator is used as the data are continuous and normally distributed and are therefore appropriate for the ML estimator. Additionally, this follows the steps in the main article and aligns with the estimator used to determine the suggested fit index cut off values (Hu and Bentler, 1999). In situations where the data are known to be nonnormally distributed the robust maximum likelihood estimator (MLR) is more appropriate and can be specified with the command `estimator=”MLR”`. The results from ML and MLR are equivalent if the data are normal, and interested readers can confirm this for themselves since `lavaan` prints the output of both ML and MLR simultaneously when MLR is used. Later sections of this ESI will describe how to modify the code to accommodate categorical data. Finally, specify that the mean structure (intercepts) should be explicitly shown.

```
STEM.step0<-cfa(data=combined.invar.mean %>% filter(group=="STEM
                Majors"), model=model.test, estimator="ML",
                meanstructure=TRUE)
```

The `summary()` function provides a convenient way to view the fit statistics and model parameters from the model that was just fit to the STEM majors data.

```
summary(STEM.step0, standardized=TRUE, fit.measures=TRUE)
```

Though the output provided by `summary()` is extensive the key fit indices are indicated by boxes in Figure ESI4. Note that the fit indices match Table 1 in the manuscript and show essentially perfect fit: CFI > 0.95; SRMR < 0.08; RMSEA < 0.06 (Hu and Bentler, 1999).

```
> STEM.step0<-cfa(data = combined.invar.mean %>% filter(group=="STEM
Majors"), model = model.test, estimator="ML", meanstructure=TRUE)
> summary(STEM.step0, standardized=TRUE, fit.measures=TRUE)
lavaan 0.6-5 ended normally after 27 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of free parameters                         39

  Number of observations                          1000

Model Test User Model:

  Test statistic                                65.438
  Degrees of freedom                                51
  P-value (Chi-square)                           0.084

Model Test Baseline Model:

  Test statistic                              6052.309
  Degrees of freedom                                66
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.998
  Tucker-Lewis Index (TLI)                       0.997

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)             -13835.349
  Loglikelihood unrestricted model (H1)     -13802.630

  Akaike (AIC)                               27748.698
  Bayesian (BIC)                             27940.100
  Sample-size adjusted Bayesian (BIC)        27816.234

Root Mean Square Error of Approximation:

  RMSEA                                          0.017
  90 Percent confidence interval - lower         0.000
  90 Percent confidence interval - upper         0.028
  P-value RMSEA <= 0.05                          1.000

Standardized Root Mean Square Residual:

  SRMR                                           0.021
```

Figure ESI4. Summary output for testing baseline model (Step 0) with STEM majors data having modified I3 intercept highlighting chi square test statistic, degrees of freedom, *p*-value, CFI, RMSEA and SRMR.

The same code can be executed using the non-STEM majors data and nearly identical fit is achieved (Figure ESI5).

```
nonSTEM.step0<-cfa(data=combined.invar.mean %>% filter(group=="Non-
                   STEM Majors"), model=model.test,
                   estimator="ML", meanstructure=TRUE)

summary(nonSTEM.step0, standardized=TRUE, fit.measures=TRUE)
```

```
> nonSTEM.step0<-cfa(data = combined.invar.mean %>% filter(group=="Non-
STEM Majors"), model = model.test, estimator="ML", meanstructure=TRUE)
> summary(nonSTEM.step0, standardized=TRUE, fit.measures=TRUE)
lavaan 0.6-5 ended normally after 30 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of free parameters                         39

  Number of observations                          1000

Model Test User Model:

  Test statistic                                51.931
  Degrees of freedom                                51
  P-value (Chi-square)                           0.437

Model Test Baseline Model:

  Test statistic                              6015.854
  Degrees of freedom                                66
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    1.000
  Tucker-Lewis Index (TLI)                       1.000

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)             -13981.961
  Loglikelihood unrestricted model (H1)     -13955.996

  Akaike (AIC)                               28041.922
  Bayesian (BIC)                             28233.325
  Sample-size adjusted Bayesian (BIC)        28109.459

Root Mean Square Error of Approximation:

  RMSEA                                          0.004
  90 Percent confidence interval - lower         0.000
  90 Percent confidence interval - upper         0.021
  P-value RMSEA <= 0.05                          1.000

Standardized Root Mean Square Residual:

  SRMR                                           0.016
```

Figure ESI5. R summary output for testing baseline model (Step 0) with unmodified non-STEM majors data highlighting chi square test statistic, degrees of freedom, *p*-value, CFI, RMSEA and SRMR.

Looking through the rest of the `summary()` output gives the values for the model parameters. The column `Std.all` is most typically reported when standardized model parameters are given. For both groups, these model parameters (Figures ESI6 & ESI7) match those used to simulate the data (loadings of 0.80 as well as associations between the three factors of approximately 0.3, 0.2, and 0.1). Examining the values of the intercept terms in both groups shows that in the STEM majors group (Figure ESI6) the intercept for I3 is larger than in the non-STEM majors group by a value of 2, as specified in the model used to simulate the data.

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  IC =~
    I1                1.000                               0.761    0.787
    I2                1.030    0.040   25.575    0.000    0.784    0.793
    I3                1.053    0.040   26.167    0.000    0.802    0.802
    I4                1.075    0.041   26.429    0.000    0.818    0.815
  CC =~
    I5                1.000                               0.774    0.793
    I6                1.017    0.039   26.278    0.000    0.788    0.805
    I7                1.001    0.039   25.971    0.000    0.775    0.796
    I8                1.011    0.039   26.213    0.000    0.783    0.801
  AC =~
    I9                1.000                               0.765    0.787
    I10               1.061    0.041   25.790    0.000    0.812    0.810
    I11               0.993    0.039   25.222    0.000    0.760    0.781
    I12               1.027    0.041   25.268    0.000    0.786    0.792

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  IC ~~
    CC                0.174    0.023    7.666    0.000    0.295    0.295
    AC                0.119    0.022    5.453    0.000    0.205    0.205
  CC ~~
    AC                0.087    0.022    3.965    0.000    0.146    0.146

Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .I1               -0.004    0.031   -0.116    0.908   -0.004   -0.004
   .I2                0.012    0.031    0.391    0.696    0.012    0.012
   .I3                1.974    0.032   62.479    0.000    1.974    1.976
   .I4               -0.003    0.032   -0.082    0.935   -0.003   -0.003
   .I5               -0.009    0.031   -0.304    0.761   -0.009   -0.010
   .I6                0.004    0.031    0.142    0.887    0.004    0.004
   .I7                0.033    0.031    1.077    0.281    0.033    0.034
   .I8                0.013    0.031    0.431    0.666    0.013    0.014
   .I9               -0.018    0.031   -0.571    0.568   -0.018   -0.018
   .I10              -0.006    0.032   -0.191    0.849   -0.006   -0.006
   .I11              -0.001    0.031   -0.026    0.979   -0.001   -0.001
   .I12              -0.021    0.031   -0.679    0.497   -0.021   -0.021
    IC                0.000                               0.000    0.000
    CC                0.000                               0.000    0.000
    AC                0.000                               0.000    0.000
```

Figure ESI6. R summary output for testing baseline model (Step 0) with unchanged STEM majors data highlighting standardized model parameters and intercepts.

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  IC =~
    I1                1.000                               0.790    0.789
    I2                1.036    0.039   26.608    0.000    0.819    0.812
    I3                0.999    0.038   26.250    0.000    0.789    0.806
    I4                1.029    0.039   26.113    0.000    0.813    0.800
  CC =~
    I5                1.000                               0.780    0.796
    I6                1.058    0.039   27.089    0.000    0.825    0.821
    I7                1.027    0.039   26.141    0.000    0.801    0.791
    I8                1.051    0.040   26.545    0.000    0.820    0.804
  AC =~
    I9                1.000                               0.735    0.759
    I10               1.074    0.045   23.957    0.000    0.789    0.796
    I11               1.042    0.044   23.861    0.000    0.766    0.780
    I12               1.037    0.044   23.641    0.000    0.762    0.783

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  IC ~~
    CC                0.206    0.024    8.565    0.000    0.335    0.335
    AC                0.145    0.022    6.542    0.000    0.250    0.250
  CC ~~
    AC                0.102    0.021    4.783    0.000    0.178    0.178

Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .I1               -0.008    0.032   -0.237    0.812   -0.008   -0.008
   .I2               -0.019    0.032   -0.602    0.547   -0.019   -0.019
   .I3               -0.028    0.031   -0.897    0.370   -0.028   -0.028
   .I4               -0.050    0.032   -1.543    0.123   -0.050   -0.049
   .I5               -0.023    0.031   -0.748    0.455   -0.023   -0.024
   .I6               -0.053    0.032   -1.669    0.095   -0.053   -0.053
   .I7               -0.036    0.032   -1.128    0.259   -0.036   -0.036
   .I8               -0.005    0.032   -0.157    0.875   -0.005   -0.005
   .I9                0.041    0.031    1.330    0.183    0.041    0.042
   .I10               0.071    0.031    2.253    0.024    0.071    0.071
   .I11               0.035    0.031    1.128    0.259    0.035    0.036
   .I12               0.048    0.031    1.571    0.116    0.048    0.050
    IC                0.000                               0.000    0.000
    CC                0.000                               0.000    0.000
    AC                0.000                               0.000    0.000
```

Figure ESI7. R summary output for testing baseline model (Step 0) with unchanged non-STEM majors data highlighting standardized model parameters and intercepts.

It is important to note that this difference in intercept for I3 between the groups (Figures ESI6 & ESI7) did not affect the overall fit of each group (Figures ESI4 & ESI5) because the parameters in each group were allowed to vary as needed to best fit the model. The purpose of testing these baseline models is to ensure that each group has a reasonable fit to the model before constraining any parameters to be equal across groups.

Step 1: Configural Invariance

The next step of invariance testing fits the model to both groups of data simultaneously. Within the `cfa()` function this is easily accomplished by specifying that groups are present and providing the name of the grouping variable (`group="group"`).

```
step1.comb.mean<-cfa(data=combined.invar.mean, model=model.test,
                     group="group", estimator="ML")

summary(step1.comb.mean, standardized=TRUE, fit.measures=TRUE)
```

Output from testing this model provides both an overall model chi square and the individual group chi square values obtained from Step 0 (Figure ESI8). The rest of the fit indices (CFI, RMSEA, and SRMR) are provided for the overall model. As show in Table 1 of the manuscript the fit indices for the configural model are essentially perfect. Further exploration of the model parameters shows that parameters for both groups have been estimated separately and match those in Step 0.

```
> step1.comb.mean<-cfa(data = combined.invar.mean, model = model.test,
group="group", estimator="ML")
> summary(step1.comb.mean, standardized=TRUE, fit.measures=TRUE)
lavaan 0.6-5 ended normally after 33 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of free parameters                         78

  Number of observations per group:
    STEM Majors                                   1000
    Non-STEM Majors                               1000

Model Test User Model:

  Test statistic                               117.369
  Degrees of freedom                               102
  P-value (Chi-square)                           0.142
  Test statistic for each group:
    STEM Majors                                 65.438
    Non-STEM Majors                             51.931

Model Test Baseline Model:

  Test statistic                             12068.162
  Degrees of freedom                               132
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.999
  Tucker-Lewis Index (TLI)                       0.998

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)             -27817.310
  Loglikelihood unrestricted model (H1)     -27758.626

  Akaike (AIC)                               55790.620
  Bayesian (BIC)                             56227.490
  Sample-size adjusted Bayesian (BIC)        55979.680

Root Mean Square Error of Approximation:

  RMSEA                                          0.012
  90 Percent confidence interval - lower         0.000
  90 Percent confidence interval - upper         0.021
  P-value RMSEA <= 0.05                          1.000

Standardized Root Mean Square Residual:

  SRMR                                           0.018
```

Figure ESI8. R summary output for configural invariance model (Step 1) with STEM majors data having modified I3 intercept highlighting chi square test statistic, degrees of freedom, *p*-value, CFI, RMSEA and SRMR.

Step 2: Metric Invariance (Weak)

To test for metric invariance (weak) the `group.equal` argument is used to specify that the loadings must be held constant across the two groups.

```
step2.comb.mean<-cfa(data=combined.invar.mean, model=model.test,
                     group="group", estimator="ML",
                     group.equal=c("loadings"))

summary(step2.comb.mean, standardized=TRUE, fit.measures=TRUE)
```

The fit indices for the metric invariance model (Figure ESI9) again match Table 1 in the manuscript and show essentially perfect fit. As described in the manuscript the change in fit index values can be calculated by hand but the *p*-value for the Δchi square must be computed.

```
> step2.comb.mean<-cfa(data = combined.invar.mean, model = model.test,
group="group", estimator="ML", group.equal=c("loadings"))
> summary(step2.comb.mean, standardized=TRUE, fit.measures=TRUE)
lavaan 0.6-5 ended normally after 30 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of free parameters                         78
  Number of equality constraints                     9
  Row rank of the constraints matrix                 9

  Number of observations per group:
    STEM Majors                                   1000
    Non-STEM Majors                               1000

Model Test User Model:

  Test statistic                               120.834
  Degrees of freedom                               111
  P-value (Chi-square)                           0.246
  Test statistic for each group:
    STEM Majors                                 67.162
    Non-STEM Majors                             53.672

Model Test Baseline Model:

  Test statistic                             12068.162
  Degrees of freedom                               132
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.999
  Tucker-Lewis Index (TLI)                       0.999

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)             -27819.043
  Loglikelihood unrestricted model (H1)     -27758.626

  Akaike (AIC)                               55776.085
  Bayesian (BIC)                             56162.547
  Sample-size adjusted Bayesian (BIC)        55943.331

Root Mean Square Error of Approximation:

  RMSEA                                          0.009
  90 Percent confidence interval - lower         0.000
  90 Percent confidence interval - upper         0.019
  P-value RMSEA <= 0.05                          1.000

Standardized Root Mean Square Residual:

  SRMR                                           0.019
```

Figure ESI9. R summary output for metric invariance model (Step 2) with STEM majors data having modified I3 intercept highlighting chi square test statistic, degrees of freedom, *p*-value, CFI, RMSEA and SRMR.

Examination of the model parameters is again done by groups (Figure ESI10) but shows that certain parameters have been constrained equal across the groups by assigning them a parameter name given in parenthesis (e.g., .p2.). Here the unstandardized loading values in the `Estimate` column are equal in both groups but the `Std.all` column values vary slightly. This is because the factors parameters (i.e., factor covariances) have not been constrained equal across groups and therefore affect the standardized loading values. Note that only the loadings have been assigned parameter names since these are the only parameters constrained equal across groups.

```
Group 1 [STEM Majors]:

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  IC =~
    I1                1.000                               0.770    0.791
    I2      (.p2.)    1.034    0.028   36.926    0.000    0.796    0.799
    I3      (.p3.)    1.025    0.028   36.999    0.000    0.789    0.796
    I4      (.p4.)    1.053    0.028   37.170    0.000    0.811    0.812
  CC =~
    I5                1.000                               0.763    0.788
    I6      (.p6.)    1.038    0.027   37.797    0.000    0.793    0.807
    I7      (.p7.)    1.015    0.028   36.792    0.000    0.775    0.796
    I8      (.p8.)    1.032    0.028   37.269    0.000    0.788    0.803
  AC =~
    I9                1.000                               0.759    0.784
    I10     (.10.)    1.067    0.030   35.415    0.000    0.810    0.809
    I11     (.11.)    1.016    0.030   34.431    0.000    0.771    0.787
    I12     (.12.)    1.031    0.030   34.755    0.000    0.783    0.790

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  IC ~~
    CC                0.174    0.022    7.785    0.000    0.295    0.295
    AC                0.119    0.022    5.501    0.000    0.204    0.204
  CC ~~
    AC                0.084    0.021    3.976    0.000    0.146    0.146

Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .I1               -0.004    0.031   -0.115    0.908   -0.004   -0.004
   .I2                0.012    0.031    0.388    0.698    0.012    0.012
   .I3                1.974    0.031   62.973    0.000    1.974    1.991
   .I4               -0.003    0.032   -0.082    0.934   -0.003   -0.003
   .I5               -0.009    0.031   -0.306    0.760   -0.009   -0.010
   .I6                0.004    0.031    0.141    0.887    0.004    0.004
   .I7                0.033    0.031    1.077    0.281    0.033    0.034
   .I8                0.013    0.031    0.430    0.667    0.013    0.014
   .I9               -0.018    0.031   -0.574    0.566   -0.018   -0.018
   .I10              -0.006    0.032   -0.191    0.849   -0.006   -0.006
   .I11              -0.001    0.031   -0.026    0.979   -0.001   -0.001
   .I12              -0.021    0.031   -0.681    0.496   -0.021   -0.022
    IC                0.000                               0.000    0.000
    CC                0.000                               0.000    0.000
    AC                0.000                               0.000    0.000
```

```
Group 2 [Non-STEM Majors]:

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  IC =~
    I1                1.000                               0.781    0.784
    I2      (.p2.)    1.034    0.028   36.926    0.000    0.807    0.806
    I3      (.p3.)    1.025    0.028   36.999    0.000    0.800    0.812
    I4      (.p4.)    1.053    0.028   37.170    0.000    0.822    0.804
  CC =~
    I5                1.000                               0.790    0.801
    I6      (.p6.)    1.038    0.027   37.797    0.000    0.820    0.819
    I7      (.p7.)    1.015    0.028   36.792    0.000    0.801    0.791
    I8      (.p8.)    1.032    0.028   37.269    0.000    0.815    0.802
  AC =~
    I9                1.000                               0.742    0.762
    I10     (.10.)    1.067    0.030   35.415    0.000    0.791    0.797
    I11     (.11.)    1.016    0.030   34.431    0.000    0.754    0.773
    I12     (.12.)    1.031    0.030   34.755    0.000    0.765    0.785

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  IC ~~
    CC                0.207    0.024    8.738    0.000    0.335    0.335
    AC                0.145    0.022    6.625    0.000    0.250    0.250
  CC ~~
    AC                0.104    0.022    4.804    0.000    0.178    0.178

Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .I1               -0.008    0.031   -0.239    0.811   -0.008   -0.008
   .I2               -0.019    0.032   -0.606    0.544   -0.019   -0.019
   .I3               -0.028    0.031   -0.891    0.373   -0.028   -0.028
   .I4               -0.050    0.032   -1.535    0.125   -0.050   -0.049
   .I5               -0.023    0.031   -0.743    0.458   -0.023   -0.023
   .I6               -0.053    0.032   -1.674    0.094   -0.053   -0.053
   .I7               -0.036    0.032   -1.128    0.259   -0.036   -0.036
   .I8               -0.005    0.032   -0.158    0.875   -0.005   -0.005
   .I9                0.041    0.031    1.324    0.185    0.041    0.042
   .I10               0.071    0.031    2.250    0.024    0.071    0.071
   .I11               0.035    0.031    1.137    0.256    0.035    0.036
   .I12               0.048    0.031    1.568    0.117    0.048    0.050
    IC                0.000                               0.000    0.000
    CC                0.000                               0.000    0.000
    AC                0.000                               0.000    0.000
```

Figure ESI10. R summary output for metric invariance model (Step 2) with STEM majors data having modified I3 intercept highlighting constraints on loading terms.

Step 3: Scalar Invariance (Strong)

Testing for scalar invariance only requires the addition of constraining the intercept terms to be equal, in addition to the loadings that were already constrained in Step 2.

```
step3.comb.mean<-cfa(data=combined.invar.mean, model=model.test,
                     group="group", estimator="ML",
                     group.equal=c("loadings", "intercepts"))

summary(step3.comb.mean, standardized=TRUE, fit.measures=TRUE)
```

Again, matching the values found in Table 1 of the manuscript, the fit indices for the strict invariance model (Figure ESI11) indicate poor data-model fit, which is to be expected since the intercept terms were not simulated to be equal across groups. Notice that the chi square values for the individual groups give some indication that the problem is in the STEM Majors group, as it has a much larger (worse) chi square value. Figure ESI12 shows that now the intercept terms are constrained to be equal across groups.

```
> step3.comb.mean<-cfa(data = combined.invar.mean, model = model.test,
group="group", estimator="ML", group.equal=c("loadings", "intercepts"))
> summary(step3.comb.mean, standardized=TRUE, fit.measures=TRUE)
lavaan 0.6-5 ended normally after 49 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of free parameters                         81
  Number of equality constraints                    21
  Row rank of the constraints matrix                21

  Number of observations per group:
    STEM Majors                                   1000
    Non-STEM Majors                               1000

Model Test User Model:

  Test statistic                              2267.834
  Degrees of freedom                               120
  P-value (Chi-square)                           0.000
  Test statistic for each group:
    STEM Majors                               2067.996
    Non-STEM Majors                            199.838

Model Test Baseline Model:

  Test statistic                             12068.162
  Degrees of freedom                               132
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.820
  Tucker-Lewis Index (TLI)                       0.802

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)             -28892.543
  Loglikelihood unrestricted model (H1)     -27758.626

  Akaike (AIC)                               57905.085
  Bayesian (BIC)                             58241.139
  Sample-size adjusted Bayesian (BIC)        58050.516

Root Mean Square Error of Approximation:

  RMSEA                                          0.134
  90 Percent confidence interval - lower         0.129
  90 Percent confidence interval - upper         0.139
  P-value RMSEA <= 0.05                          0.000

Standardized Root Mean Square Residual:

  SRMR                                           0.191
```

Figure ESI11. R summary output for metric invariance model (Step 3) with STEM majors data having modified I3 intercept highlighting chi square test statistic, degrees of freedom, *p*-value, CFI, RMSEA and SRMR.

```
Group 1 [STEM Majors]:

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  IC =~
    I1                1.000                               0.759    0.781
    I2      (.p2.)    1.043    0.030   35.235    0.000    0.791    0.798
    I3      (.p3.)    1.044    0.036   29.207    0.000    0.793    0.402
    I4      (.p4.)    1.068    0.030   35.439    0.000    0.810    0.815
  CC =~
    I5                1.000                               0.763    0.788
    I6      (.p6.)    1.039    0.027   37.804    0.000    0.793    0.807
    I7      (.p7.)    1.015    0.028   36.786    0.000    0.775    0.796
    I8      (.p8.)    1.032    0.028   37.252    0.000    0.787    0.803
  AC =~
    I9                1.000                               0.759    0.784
    I10     (.10.)    1.067    0.030   35.442    0.000    0.810    0.809
    I11     (.11.)    1.015    0.029   34.442    0.000    0.771    0.787
    I12     (.12.)    1.032    0.030   34.788    0.000    0.783    0.790

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  IC ~~
    CC                0.170    0.022    7.612    0.000    0.294    0.294
    AC                0.108    0.022    4.973    0.000    0.187    0.187
  CC ~~
    AC                0.084    0.021    3.976    0.000    0.146    0.146

Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .I1      (.31.)    0.066    0.029    2.296    0.022    0.066    0.067
   .I2      (.32.)    0.073    0.029    2.480    0.013    0.073    0.073
   .I3      (.33.)    0.324    0.034    9.449    0.000    0.324    0.164
   .I4      (.34.)    0.049    0.030    1.641    0.101    0.049    0.049
   .I5      (.35.)    0.003    0.028    0.119    0.905    0.003    0.003
   .I6      (.36.)   -0.004    0.029   -0.141    0.888   -0.004   -0.004
   .I7      (.37.)    0.019    0.029    0.665    0.506    0.019    0.020
   .I8      (.38.)    0.024    0.029    0.823    0.410    0.024    0.024
   .I9      (.39.)   -0.018    0.028   -0.629    0.529   -0.018   -0.018
   .I10     (.40.)    0.001    0.030    0.024    0.981    0.001    0.001
   .I11     (.41.)   -0.013    0.029   -0.436    0.663   -0.013   -0.013
   .I12     (.42.)   -0.017    0.029   -0.581    0.561   -0.017   -0.017
    IC                0.000                               0.000    0.000
    CC                0.000                               0.000    0.000
    AC                0.000                               0.000    0.000
```

```
Group 2 [Non-STEM Majors]:

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  IC =~
    I1                1.000                               0.770    0.777
    I2      (.p2.)    1.043    0.030   35.235    0.000    0.803    0.804
    I3      (.p3.)    1.044    0.036   29.207    0.000    0.805    0.789
    I4      (.p4.)    1.068    0.030   35.439    0.000    0.822    0.804
  CC =~
    I5                1.000                               0.789    0.800
    I6      (.p6.)    1.039    0.027   37.804    0.000    0.820    0.819
    I7      (.p7.)    1.015    0.028   36.786    0.000    0.802    0.791
    I8      (.p8.)    1.032    0.028   37.252    0.000    0.814    0.802
  AC =~
    I9                1.000                               0.742    0.762
    I10     (.10.)    1.067    0.030   35.442    0.000    0.792    0.797
    I11     (.11.)    1.015    0.029   34.442    0.000    0.753    0.773
    I12     (.12.)    1.032    0.030   34.788    0.000    0.765    0.785

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  IC ~~
    CC                0.205    0.023    8.726    0.000    0.337    0.337
    AC                0.144    0.022    6.636    0.000    0.252    0.252
  CC ~~
    AC                0.104    0.022    4.805    0.000    0.178    0.178

Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .I1      (.31.)    0.066    0.029    2.296    0.022    0.066    0.066
   .I2      (.32.)    0.073    0.029    2.480    0.013    0.073    0.073
   .I3      (.33.)    0.324    0.034    9.449    0.000    0.324    0.318
   .I4      (.34.)    0.049    0.030    1.641    0.101    0.049    0.048
   .I5      (.35.)    0.003    0.028    0.119    0.905    0.003    0.003
   .I6      (.36.)   -0.004    0.029   -0.141    0.888   -0.004   -0.004
   .I7      (.37.)    0.019    0.029    0.665    0.506    0.019    0.019
   .I8      (.38.)    0.024    0.029    0.823    0.410    0.024    0.023
   .I9      (.39.)   -0.018    0.028   -0.629    0.529   -0.018   -0.018
   .I10     (.40.)    0.001    0.030    0.024    0.981    0.001    0.001
   .I11     (.41.)   -0.013    0.029   -0.436    0.663   -0.013   -0.013
   .I12     (.42.)   -0.017    0.029   -0.581    0.561   -0.017   -0.017
    IC               -0.146    0.037   -3.922    0.000   -0.189   -0.189
    CC               -0.039    0.037   -1.053    0.292   -0.049   -0.049
    AC                0.059    0.036    1.634    0.102    0.079    0.079
```

Figure ESI12. R summary output for scalar invariance model (Step 3) with STEM majors data having modified I3 intercept highlighting constraints on loading and intercept terms.

Step 4: Conservative Invariance (Strict)

Given the poor fit of the scalar invariance model, and out of range delta fit index values, it is not appropriate to go on to consider the strict invariance model. However, interested readers can test this model by adding `"residuals"` to the `group.equal` argument (residuals is another name for the error variance terms).

```
Step4.comb.mean<-cfa(data=combined.invar.mean, model=model.test,
                     group="group", estimator="ML",
                     group.equal=c("loadings", "intercepts", "residuals"))

summary(step4.comb.mean, standardized=TRUE, fit.measures=TRUE)
```

Exporting Data from R to Mplus

Data within R can be exported in a variety of familiar formats including txt, csv, and xlsx. Most conveniently for those working in Mplus there is also a package, `MplusAutomation` (Hallquist and Wiley, 2018), that allows for direct export of data in the correct Mplus format, dat. The correct format for Mplus requires data to not have any header information, such as column names. The `MplusAutomation` package also generates appropriate code to communicate the structure of the file to Mplus. The R code below shows how to export the simulated PRCQ data to Mplus and request the input file, which provides the code to use within Mplus to import the dat file in the correct format to be read by Mplus. Note that the group variable had been stored as a categorical factor within R and must be changed to a numeric variable for export. In this case the first group (STEM majors) will become 1 and the second group will become 2. This can be confirmed with the `describeBy()` function.

```
library(MplusAutomation)

combined.invar.mean$group<-combined.invar.mean$group %>%
                           as.numeric()
describeBy(combined.invar.mean, group="group")

prepareMplusData(combined.invar.mean,
                 filename="InvarianceMean.dat", inpfile = TRUE,
                 keepCols=c("I1", "I2", "I3", "I4","I5", "I6",
                 "I7", "I8", "I9", "I10", "I11", "I12", "group"))
```

As a result of these commands R will create two new files, `InvarianceMean.dat` and `InvarianceMean.inp` in the working directory of your R session. If you are unsure of where your working directory resides, use the command `getwd()`.

Invariance Testing with Mplus – Continuous Data

Invariance testing in Mplus begins by opening the `inp` file generated previously or creating a new `inp` file for your own data. At the top of the `inp` file will be a title for the model being tested, the name of the data file, and the names of the variables in the data file. As before, the first step should be to test the model for each group individually. This is accomplished with the command `USEOBSERVATIONS`. Then the model to be tested is specified, this step is similar to lavaan but uses the term `BY` instead of =~ to denote relations between items and factors.

```
TITLE: STEM Majors Group Step 0
DATA: FILE = "InvarianceMean.dat";
VARIABLE:
NAMES = I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12 group;
USEVARIABLES ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
USEOBSERVATIONS are group==1;

MODEL:
IC BY I1 I2 I3 I4;
CC BY I5 I6 I7 I8;
AC BY I9 I10 I11 I12;

OUTPUT:
STANDARDIZED;
```

The output for this model provides the same fit indices and standardized model parameters (Figure ESI13) as produced in R (Figures ESI4 & ESI 6) and shown in Table 1 of the manuscript.

```
MODEL FIT INFORMATION

Number of Free Parameters                       39

Loglikelihood

          H0 Value                      -13835.349
          H1 Value                      -13802.630

Information Criteria

          Akaike (AIC)                   27748.698
          Bayesian (BIC)                 27940.100
          Sample-Size Adjusted BIC       27816.234
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                             65.438
          Degrees of Freedom                    51
          P-Value                           0.0841

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.017
          90 Percent C.I.                    0.000  0.028
          Probability RMSEA <= .05           1.000

CFI/TLI

          CFI                                0.998
          TLI                                0.997

Chi-Square Test of Model Fit for the Baseline Model

          Value                           6052.309
          Degrees of Freedom                    66
          P-Value                           0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                              0.021
```

```
STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 IC       BY
    I1                 0.787      0.015     52.142      0.000
    I2                 0.793      0.015     53.520      0.000
    I3                 0.802      0.014     55.343      0.000
    I4                 0.815      0.014     58.397      0.000

 CC       BY
    I5                 0.793      0.015     53.401      0.000
    I6                 0.805      0.014     55.833      0.000
    I7                 0.796      0.015     53.890      0.000
    I8                 0.801      0.015     54.917      0.000

 AC       BY
    I9                 0.787      0.015     51.270      0.000
    I10                0.810      0.014     55.917      0.000
    I11                0.781      0.016     50.057      0.000
    I12                0.792      0.015     52.257      0.000

 CC       WITH
    IC                 0.295      0.033      8.851      0.000

 AC       WITH
    IC                 0.205      0.035      5.866      0.000
    CC                 0.146      0.036      4.107      0.000

 Intercepts
    I1                -0.004      0.032     -0.116      0.908
    I2                 0.012      0.032      0.391      0.696
    I3                 1.976      0.054     36.366      0.000
    I4                -0.003      0.032     -0.082      0.935
    I5                -0.010      0.032     -0.304      0.761
    I6                 0.004      0.032      0.142      0.887
    I7                 0.034      0.032      1.077      0.282
    I8                 0.014      0.032      0.431      0.666
    I9                -0.018      0.032     -0.571      0.568
    I10               -0.006      0.032     -0.191      0.849
    I11               -0.001      0.032     -0.026      0.979
    I12               -0.021      0.032     -0.679      0.497
```

Figure ESI13. Mplus summary output baseline model (Step 0) with STEM majors data having modified I3 intercept highlighting chi square test statistic, degrees of freedom, *p*-value, CFI, RMSEA, SRMR, and standardized model parameters.

Similar code can be used for the non-STEM majors group and again the results (Figure ESI14) will agree with the R output (Figures ESI15 & ESI17 as well as Table 1 of the manuscript.

```
TITLE: Non-STEM Majors Group Step 0
DATA: FILE = "InvarianceMean.dat";
VARIABLE:
NAMES = I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12 group;
USEVARIABLES ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
USEOBSERVATIONS are group==2;

MODEL:
IC BY I1 I2 I3 I4;
CC BY I5 I6 I7 I8;
AC BY I9 I10 I11 I12;

OUTPUT:
STANDARDIZED;
```

```
MODEL FIT INFORMATION

Number of Free Parameters                        39

Loglikelihood

          H0 Value                       -13981.961
          H1 Value                       -13955.996

Information Criteria

          Akaike (AIC)                    28041.922
          Bayesian (BIC)                  28233.325
          Sample-Size Adjusted BIC        28109.459
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                              51.931
          Degrees of Freedom                     51
          P-Value                            0.4374

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                            0.004
          90 Percent C.I.                     0.000  0.021
          Probability RMSEA <= .05            1.000

CFI/TLI

          CFI                                 1.000
          TLI                                 1.000

Chi-Square Test of Model Fit for the Baseline Model

          Value                            6015.854
          Degrees of Freedom                     66
          P-Value                            0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                               0.016
```

```
STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 IC       BY
    I1                 0.789      0.015     52.836      0.000
    I2                 0.812      0.014     58.082      0.000
    I3                 0.806      0.014     56.771      0.000
    I4                 0.800      0.015     55.114      0.000

 CC       BY
    I5                 0.796      0.015     54.388      0.000
    I6                 0.821      0.014     60.409      0.000
    I7                 0.791      0.015     53.320      0.000
    I8                 0.804      0.014     56.398      0.000

 AC       BY
    I9                 0.759      0.017     44.993      0.000
    I10                0.796      0.015     51.583      0.000
    I11                0.780      0.016     48.554      0.000
    I12                0.783      0.016     49.300      0.000

 CC       WITH
    IC                 0.335      0.032     10.308      0.000

 AC       WITH
    IC                 0.250      0.034      7.268      0.000
    CC                 0.178      0.035      5.041      0.000

 Intercepts
    I1                -0.008      0.032     -0.237      0.812
    I2                -0.019      0.032     -0.602      0.547
    I3                -0.028      0.032     -0.897      0.370
    I4                -0.049      0.032     -1.542      0.123
    I5                -0.024      0.032     -0.748      0.455
    I6                -0.053      0.032     -1.668      0.095
    I7                -0.036      0.032     -1.128      0.259
    I8                -0.005      0.032     -0.157      0.875
    I9                 0.042      0.032      1.330      0.184
    I10                0.071      0.032      2.250      0.024
    I11                0.036      0.032      1.128      0.259
    I12                0.050      0.032      1.570      0.116
```

Figure ESI14. Mplus summary output baseline model (Step 0) with Non-STEM majors data having modified I3 intercept highlighting chi square test statistic, degrees of freedom, *p*-value, CFI, RMSEA, SRMR, and standardized model parameters.

Step 1: Configural Invariance

To test configural invariance within Mplus, the model is specified separately for each group. The `!` notation is used to insert comments within the Mplus model code. To provide results aligned with the R output the `@1` notation is used to identify the model by standardizing the loading for the first item on each factor. This is the default setting for the R `cfa()` function, but models in both programs can also be run by standardizing the factors instead of the loadings as a method of identifying the model.

Next the factor intercept is set to zero using brackets and `@0` notation. By default, Mplus assumes that item intercepts should be equal across groups, these can be freely estimated using the bracket notation. Item error variances are coded without the use of brackets. Specifying the same model for the second group will tell Mplus to estimate parameters for both models separately.

```
TITLE: Combined Dataset with Mean Differences Step 1 (Configural)
DATA: FILE = "InvarianceMean.dat";
VARIABLE:
NAMES = I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12 group;
USEVARIABLES ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
GROUPING = group (1 = STEM 2 = NonSTEM);

MODEL:
! Model with standardized loading of first item on each factor
  IC BY I1@1 I2 I3 I4;
  CC BY I5@1 I6 I7 I8;
  AC BY I9@1 I10 I11 I12;

! Setting factor intercepts to zero
  [IC@0];
  [CC@0];
  [AC@0];

! Allowing item intercepts to be freely estimated
  [I1-I12];

! Allowing item error variances to be freely estimated
  I1-I12;

! Specifying the same model for the second group will cause
! all parameters to be freely estimated for the second group
MODEL NonSTEM:
  IC BY I1@1 I2 I3 I4;
  CC BY I5@1 I6 I7 I8;
  AC BY I9@1 I10 I11 I12;

  [IC@0];
  [CC@0];
  [AC@0];

  [I1-I12];

  I1-I12;
OUTPUT:
STANDARDIZED;
```

The output from this model (Figure ESI15) matches the fit indices in Table 1 of the manuscript for the configural model and both the unstandardized and standardized model parameters for the STEM majors group (Figure ESI16) and non-STEM majors group match those found using R (Figures ESI6 & ESI7).

```
MODEL FIT INFORMATION

Number of Free Parameters                         78

Loglikelihood

          H0 Value                        -27817.310
          H1 Value                        -27758.626

Information Criteria

          Akaike (AIC)                     55790.620
          Bayesian (BIC)                   56227.490
          Sample-Size Adjusted BIC         55979.680
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                              117.369
          Degrees of Freedom                     102
          P-Value                             0.1418

Chi-Square Contribution From Each Group

          STEM                                65.438
          NONSTEM                             51.931

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                             0.012
          90 Percent C.I.                      0.000  0.021
          Probability RMSEA <= .05             1.000

CFI/TLI

          CFI                                  0.999
          TLI                                  0.998

Chi-Square Test of Model Fit for the Baseline Model

          Value                            12068.162
          Degrees of Freedom                     132
          P-Value                             0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                                0.019
```

Figure ESI15. Mplus summary output for configural invariance (Step 1) with STEM majors data having modified I3 intercept highlighting fit information.

```
MODEL RESULTS

                    Estimate

Group STEM

 IC       BY
    I1                 1.000
    I2                 1.030
    I3                 1.053
    I4                 1.075

 CC       BY
    I5                 1.000
    I6                 1.017
    I7                 1.001
    I8                 1.011

 AC       BY
    I9                 1.000
    I10                1.061
    I11                0.993
    I12                1.027

 CC       WITH
    IC                 0.174

 AC       WITH
    IC                 0.119
    CC                 0.086

 Means
    IC                 0.000
    CC                 0.000
    AC                 0.000

 Intercepts
    I1                -0.004
    I2                 0.012
    I3                 1.974
    I4                -0.003
    I5                -0.009
    I6                 0.004
    I7                 0.033
    I8                 0.013
    I9                -0.018
    I10               -0.006
    I11               -0.001
    I12               -0.021
```

```
STDYX Standardization

                    Estimate

Group STEM

 IC       BY
    I1                 0.787
    I2                 0.793
    I3                 0.802
    I4                 0.815

 CC       BY
    I5                 0.793
    I6                 0.805
    I7                 0.796
    I8                 0.801

 AC       BY
    I9                 0.787
    I10                0.810
    I11                0.781
    I12                0.792

 CC       WITH
    IC                 0.295

 AC       WITH
    IC                 0.205
    CC                 0.146

 Means
    IC                 0.000
    CC                 0.000
    AC                 0.000

 Intercepts
    I1                -0.004
    I2                 0.012
    I3                 1.976
    I4                -0.003
    I5                -0.010
    I6                 0.004
    I7                 0.034
    I8                 0.014
    I9                -0.018
    I10               -0.006
    I11               -0.001
    I12               -0.021
```

```
MODEL RESULTS

Group NONSTEM

 IC       BY
    I1                 1.000
    I2                 1.036
    I3                 0.999
    I4                 1.029

 CC       BY
    I5                 1.000
    I6                 1.058
    I7                 1.027
    I8                 1.051

 AC       BY
    I9                 1.000
    I10                1.074
    I11                1.043
    I12                1.037

 CC       WITH
    IC                 0.206

 AC       WITH
    IC                 0.145
    CC                 0.102

 Means
    IC                 0.000
    CC                 0.000
    AC                 0.000

 Intercepts
    I1                -0.008
    I2                -0.019
    I3                -0.028
    I4                -0.050
    I5                -0.023
    I6                -0.053
    I7                -0.036
    I8                -0.005
    I9                 0.041
    I10                0.071
    I11                0.035
    I12                0.048
```

```
STDYX Standardization

Group NONSTEM

 IC       BY
    I1                 0.789
    I2                 0.812
    I3                 0.806
    I4                 0.800

 CC       BY
    I5                 0.796
    I6                 0.821
    I7                 0.791
    I8                 0.804

 AC       BY
    I9                 0.759
    I10                0.796
    I11                0.780
    I12                0.783

 CC       WITH
    IC                 0.335

 AC       WITH
    IC                 0.250
    CC                 0.178

 Means
    IC                 0.000
    CC                 0.000
    AC                 0.000

 Intercepts
    I1                -0.008
    I2                -0.019
    I3                -0.028
    I4                -0.049
    I5                -0.024
    I6                -0.053
    I7                -0.036
    I8                -0.005
    I9                 0.042
    I10                0.071
    I11                0.036
    I12                0.050
```

Figure ESI16. Mplus output for configural invariance (Step 1) with STEM majors data having modified I3 intercept highlighting unstandardized and standardized model parameters for both groups.

Step 2: Metric Invariance (Weak)

Metric invariance is tested by assigning the same parameter names to the loading terms in each group. In this example the names `L1-L12` are assigned to each of the loading parameters. Repeating this assignment in the second group will cause Mplus to set the unstandardized value of the parameters equal.

```
TITLE: Combined Dataset with Mean Differences Step 2 (Weak)
DATA: FILE = "InvarianceMean.dat";
VARIABLE:
NAMES = I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12 group;
USEVARIABLES ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
GROUPING = group (1 = STEM 2 = NonSTEM);

MODEL:
! Model with standardized loading of first item on each factor
! Assigning a parameter name to each loading value (L1-L12)
  IC BY I1@1 I2 I3 I4 (L1-L4);
  CC BY I5@1 I6 I7 I8 (L5-L8);
  AC BY I9@1 I10 I11 I12 (L9-L12);

! Setting factor intercepts to zero
  [IC@0];
  [CC@0];
  [AC@0];

! Allowing item intercepts to be freely estimated
  [I1-I12];

! Allowing item error variances to be freely estimated
  I1-I12;

! Specifying the same model for the second group will force
! loadings to be equivalent across groups while other
! parameters are freely estimated
MODEL NonSTEM:
  IC BY I1@1 I2 I3 I4 (L1-L4);
  CC BY I5@1 I6 I7 I8 (L5-L8);
  AC BY I9@1 I10 I11 I12 (L9-L12);

  [IC@0];
  [CC@0];
  [AC@0];

  [I1-I12];

  I1-I12;

OUTPUT:
STANDARDIZED;
```

The output from this model (Figure ESI17) matches the fit indices in Table 1 of the manuscript for the weak invariance model and now the unstandardized parameters are equal across groups (Figure ESI18) while the intercepts are allowed to differ. As before, the standardized parameters differ slightly, but are aligned with the R output (Figure ESI10).

```
MODEL FIT INFORMATION

Number of Free Parameters                   69

Loglikelihood

          H0 Value                  -27819.043
          H1 Value                  -27758.626

Information Criteria

          Akaike (AIC)               55776.085
          Bayesian (BIC)             56162.547
          Sample-Size Adjusted BIC   55943.331
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                        120.834
          Degrees of Freedom               111
          P-Value                       0.2464

Chi-Square Contribution From Each Group

          STEM                          67.159
          NONSTEM                       53.674

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                       0.009
          90 Percent C.I.                0.000  0.019
          Probability RMSEA <= .05       1.000

CFI/TLI

          CFI                            0.999
          TLI                            0.999

Chi-Square Test of Model Fit for the Baseline Model

          Value                      12068.162
          Degrees of Freedom               132
          P-Value                       0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                          0.019
```

Figure ESI17. Mplus summary output for metric invariance (Step 2) with STEM majors data having modified I3 intercept highlighting fit information.

```
MODEL RESULTS                    STDYX Standardization

                   Estimate                        Estimate

Group STEM                       Group STEM

 IC       BY                      IC       BY
    I1               1.000           I1               0.791
    I2               1.034           I2               0.799
    I3               1.025           I3               0.796
    I4               1.053           I4               0.812

 CC       BY                      CC       BY
    I5               1.000           I5               0.788
    I6               1.038           I6               0.807
    I7               1.015           I7               0.796
    I8               1.032           I8               0.803

 AC       BY                      AC       BY
    I9               1.000           I9               0.784
    I10              1.067           I10              0.809
    I11              1.016           I11              0.787
    I12              1.031           I12              0.790

 CC       WITH                    CC       WITH
    IC               0.174           IC               0.295

 AC       WITH                    AC       WITH
    IC               0.119           IC               0.204
    CC               0.085           CC               0.146

 Means                            Means
    IC               0.000           IC               0.000
    CC               0.000           CC               0.000
    AC               0.000           AC               0.000

 Intercepts                       Intercepts
    I1              -0.004           I1              -0.004
    I2               0.012           I2               0.012
    I3               1.974           I3               1.991
    I4              -0.003           I4              -0.003
    I5              -0.009           I5              -0.010
    I6               0.004           I6               0.004
    I7               0.033           I7               0.034
    I8               0.013           I8               0.014
    I9              -0.018           I9              -0.018
    I10             -0.006           I10             -0.006
    I11             -0.001           I11             -0.001
    I12             -0.021           I12             -0.022
```

```
MODEL RESULTS                    STDYX Standardization

Group NONSTEM                    Group NONSTEM

 IC       BY                      IC       BY
    I1               1.000           I1               0.784
    I2               1.034           I2               0.806
    I3               1.025           I3               0.812
    I4               1.053           I4               0.804

 CC       BY                      CC       BY
    I5               1.000           I5               0.801
    I6               1.038           I6               0.819
    I7               1.015           I7               0.791
    I8               1.032           I8               0.802

 AC       BY                      AC       BY
    I9               1.000           I9               0.762
    I10              1.067           I10              0.797
    I11              1.016           I11              0.773
    I12              1.031           I12              0.785

 CC       WITH                    CC       WITH
    IC               0.207           IC               0.335

 AC       WITH                    AC       WITH
    IC               0.145           IC               0.250
    CC               0.104           CC               0.178

 Means                            Means
    IC               0.000           IC               0.000
    CC               0.000           CC               0.000
    AC               0.000           AC               0.000

 Intercepts                       Intercepts
    I1              -0.008           I1              -0.008
    I2              -0.019           I2              -0.019
    I3              -0.028           I3              -0.028
    I4              -0.050           I4              -0.049
    I5              -0.023           I5              -0.023
    I6              -0.053           I6              -0.053
    I7              -0.036           I7              -0.036
    I8              -0.005           I8              -0.005
    I9               0.041           I9               0.042
    I10              0.071           I10              0.071
    I11              0.035           I11              0.036
    I12              0.048           I12              0.050
```

Figure ESI18. Mplus output for metric invariance (Step 2) with STEM majors data having modified I3 intercept highlighting unstandardized and standardized model parameters for both groups.

Step 3: Scalar Invariance (Strong)

Scalar invariance is tested by assigning the same parameter names to the intercept terms in both groups while also removing the restrictions on the mean of the factor terms for the second group using the * notation. As seen in Table 1 of the manuscript and in the R output, this significantly worsens the value of all fit indices (Figure ESI19) indicating that scalar invariance has not been achieved due to differences in loadings across groups. As before, the Mplus model parameters (Figure ESI20) are similar to those produced by R (Figure ESI12).

```
TITLE: Combined Dataset with Mean Differences Step 3 (Strong)
DATA: FILE = "InvarianceMean.dat";
VARIABLE:
NAMES = I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12 group;
USEVARIABLES ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
GROUPING = group (1 = STEM 2 = NonSTEM);

MODEL:
! Model with standardized loading of first item on each factor
! Assigning a parameter name to each loading value (L1-12)
  IC BY I1@1 I2 I3 I4 (L1-L4);
  CC BY I5@1 I6 I7 I8 (L5-L8);
  AC BY I9@1 I10 I11 I12 (L9-L12);

! Setting factor intercepts to zero
  [IC@0];
  [CC@0];
  [AC@0];

! Allowing item intercepts to be freely estimated in one group
! assigning a parameter name so they will be equal across groups
  [I1-I12] (M1-M12);

! Allowing item error variances to be freely estimated
  I1-I12;

! Specifying the same model parameter names for the second group
! will cause loadings and item intercepts to be equivalent across
! groups while other parameters are freely estimated
MODEL NonSTEM:
  IC BY I1@1 I2 I3 I4 (L1-L4);
  CC BY I5@1 I6 I7 I8 (L5-L8);
  AC BY I9@1 I10 I11 I12 (L9-L12);

! Allowing factor intercepts vary
  [IC*];
  [CC*];
  [AC*];

  [I1-I12] (M1-M12);

  I1-I12;

OUTPUT:
STANDARDIZED;
```

```
MODEL FIT INFORMATION

Number of Free Parameters                       60

Loglikelihood

          H0 Value                      -28912.498
          H1 Value                      -27758.626

Information Criteria

          Akaike (AIC)                   57944.997
          Bayesian (BIC)                 58281.051
          Sample-Size Adjusted BIC       58090.428
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                           2307.745
          Degrees of Freedom                   120
          P-Value                           0.0000

Chi-Square Contribution From Each Group

          STEM                             229.366
          NONSTEM                         2078.380

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.135
          90 Percent C.I.                    0.130  0.140
          Probability RMSEA <= .05           0.000

CFI/TLI

          CFI                                0.817
          TLI                                0.798

Chi-Square Test of Model Fit for the Baseline Model

          Value                          12068.162
          Degrees of Freedom                   132
          P-Value                           0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                              0.238
```

Figure ESI19. Mplus summary output for scalar invariance (Step 3) with STEM majors data having modified I3 intercept highlighting fit information.

	MODEL RESULTS Group STEM Estimate	STDYX Standardization Group STEM Estimate	MODEL RESULTS Group NONSTEM	STDYX Standardization Group NONSTEM
IC BY				
I1	1.000	0.784	1.000	0.782
I2	1.032	0.793	1.032	0.801
I3	1.070	0.779	1.070	0.426
I4	1.058	0.808	1.058	0.806
CC BY				
I5	1.000	0.788	1.000	0.800
I6	1.039	0.807	1.039	0.819
I7	1.015	0.796	1.015	0.791
I8	1.031	0.803	1.031	0.802
AC BY				
I9	1.000	0.784	1.000	0.763
I10	1.067	0.809	1.067	0.797
I11	1.015	0.786	1.015	0.773
I12	1.032	0.790	1.032	0.785
CC WITH				
IC	0.172	0.296	0.207	0.337
AC WITH				
IC	0.117	0.203	0.149	0.258
CC	0.084	0.146	0.104	0.178
Means				
IC	0.000	0.000	-0.153	-0.197
CC	0.000	0.000	-0.039	-0.049
AC	0.000	0.000	0.059	0.079
Intercepts				
I1	0.069	0.071	0.069	0.069
I2	0.076	0.077	0.076	0.076
I3	1.754	1.682	1.754	0.897
I4	0.053	0.053	0.053	0.052
I5	0.003	0.003	0.003	0.003
I6	-0.004	-0.004	-0.004	-0.004
I7	0.019	0.020	0.019	0.019
I8	0.024	0.024	0.024	0.023
I9	-0.018	-0.018	-0.018	-0.018
I10	0.001	0.001	0.001	0.001
I11	-0.013	-0.013	-0.013	-0.013
I12	-0.017	-0.017	-0.017	-0.017

Figure ESI20. Mplus output for scalar invariance (Step 3) with STEM majors data having modified I3 intercept highlighting unstandardized and standardized model parameters for both groups.

Step 4: Conservative Invariance (Strict)

As noted previously, due to the poor fit of the scalar invariance model, you would stop at Step 3 and not go on to test Step 4 (conservative invariance with equal error variance terms). However, interested readers can test Step 4 in Mplus by providing the same name to the error variance parameters in both groups.

```
TITLE: Combined Dataset with Mean Differences Step 4 (Strict)
DATA: FILE = "InvarianceMean.dat";
VARIABLE:
NAMES = I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12 group;
USEVARIABLES ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
GROUPING = group (1 = STEM 2 = NonSTEM);

MODEL:
! Model with standardized loading of first item on each factor
! Assigning a parameter name to each loading value (L1-12)
  IC BY I1@1 I2 I3 I4 (L1-L4);
  CC BY I5@1 I6 I7 I8 (L5-L8);
  AC BY I9@1 I10 I11 I12 (L9-L12);

! Setting factor intercepts to zero
  [IC@0];
  [CC@0];
  [AC@0];

! Allow item intercepts to be freely estimated in one group but
! assigning a parameter name so they will be equal across groups
  [I1-I12] (M1-M12);

! Allow item error variances to be freely estimated but
! assigning a parameter name so they will be equal across groups
  I1-I12 (E1-E12);

! Specifying the same model parameter names for the second group
! will cause loadings and item intercepts to be equivalent across
! groups while other parameters are freely estimated
MODEL NonSTEM:
  IC BY I1@1 I2 I3 I4 (L1-L4);
  CC BY I5@1 I6 I7 I8 (L5-L8);
  AC BY I9@1 I10 I11 I12 (L9-L12);

! Allowing factor intercepts vary
  [IC*];
  [CC*];
  [AC*];

  [I1-I12] (M1-M12);

  I1-I12(E1-E12);

OUTPUT:
STANDARDIZED;
```

Fit Indices for other Continuous Datasets

Tables ESI1 & ESI2 show the data-model fit output from R produced from following the previous steps with the two other continuous datasets: `combined` and `combined.invar.load`.

Table ESI1. Measurement Invariance Testing for the PRCQ Instrument Comparing STEM Majors and Non-STEM Majors With `combined` Simulated Data for Illustration

Step	Testing level	χ2	*df*	*p*-value	CFI	SRMR	RMSEA	Δχ2	Δ*df*	*p*-value	ΔCFI	ΔSRMR	ΔRMSEA
0	STEM majors Baseline	65	51	0.084	0.998	0.021	0.017	-	-	-	-	-	-
0	Non-STEM majors Baseline	52	51	0.437	1.000	0.016	0.004	-	-	-	-	-	-
1	Configural	117	102	0.142	0.999	0.018	0.012	-	-	-	-	-	-
2	Metric	120	111	0.245	0.999	0.019	0.009	3	9	0.964	0.000	0.001	0.003
3	Scalar	127	120	0.311	0.999	0.020	0.008	7	9	0.637	0.000	0.001	0.001
4	Conservative	135	132	0.417	1.000	0.020	0.005	8	12	0.786	0.001	0.000	0.003

Note. STEM majors n = 1000. Non-STEM majors n = 1000. Simulated data was used and altered at the scalar level (intercepts) for illustrative purposes; fit indices are from R.

Table ESI2. Measurement Invariance Testing for the PRCQ Instrument Comparing STEM Majors and Non-STEM Majors With `combined.invar.load` Simulated Data for Illustration

Step	Testing level	χ2	*df*	*p*-value	CFI	SRMR	RMSEA	Δχ2	Δ*df*	*p*-value	ΔCFI	ΔSRMR	ΔRMSEA
0	STEM majors Baseline	65	51	0.084	0.998	0.021	0.017	-	-	-	-	-	-
0	Non-STEM majors Baseline	66	51	0.081	0.997	0.017	0.017	-	-	-	-	-	-
1	Configural	131	102	0.028	0.997	0.019	0.017	-	-	-	-	-	-
2	Metric	305	111	< 0.001	0.983	0.051	0.042	101	9	< 0.001	0.014	0.032	0.025
3	Scalar	310	120	< 0.001	0.984	0.051	0.040	5	9	0.834	0.001	0.000	0.002
4	Conservative	433	132	< 0.001	0.974	0.043	0.048	123	12	< 0.001	0.010	0.008	0.008

Note. STEM majors n = 1000. Non-STEM majors n = 1000. Simulated data was used and altered at the scalar level (intercepts) for illustrative purposes; fit indices are from R.

Creating Ordered Categorical Data in R

As seen in the previous examples, the data simulation function in R creates continuous data which may not be representative of data collected from instruments used in chemistry education research, which often have five-point Likert-type scales. The code below is used to take the original simulated datasets and turn them into Likert-type data by collapsing the full ranges of data for each item into five bins using the `cut()` function. Note that this process of creating categorical data from continuous data ensures that each bin will be populated, but issues with testing models can arise if authentic categorical data are collected with empty bins (e.g., no responses in the 1 category).

```
STEM.ord<-STEM
for(i in 1:12){
  var[i]<-paste0("I", i)
  STEM.ord[[var[i]]]<-as.numeric(cut(STEM[[var[i]]], breaks=5))
                  }

nonSTEM.ord<-nonSTEM
for(i in 1:12){
  var[i]<-paste0("I", i)
  nonSTEM.ord[[var[i]]]<-as.numeric(cut(nonSTEM[[var[i]]],
breaks=5))
                  }
combined.ord<-rbind(STEM.ord, nonSTEM.ord)


nonSTEM.invar.load.ord<-nonSTEM.invar.load
for(i in 1:12){
  var[i]<-paste0("I", i)
  nonSTEM.invar.load.ord[[var[i]]]<-
as.numeric(cut(nonSTEM.invar.load[[var[i]]], breaks=5))
}
combined.invar.load.ord<-rbind(STEM.ord, nonSTEM.invar.load.ord)

STEM.invar.mean.ord<-STEM.invar.mean
for(i in 1:12){
  var[i]<-paste0("I", i)
  STEM.invar.mean.ord[[var[i]]]<-
as.numeric(cut(STEM.invar.mean[[var[i]]], breaks=5))
}
combined.invar.mean.ord<-rbind(STEM.invar.mean.ord, nonSTEM.ord)
```

When data collected on Likert-type scales have fewer than seven categories or the full range of the response scale is not used by most respondents (i.e. a ceiling or floor effect) it is often recommended to treat the data as ordinal categorical data rather than continuous. In a factor analysis framework, this type of data is best modeled using a robust diagonally weighted least squares estimator, such as WLSMV (Finney and DiStefano, 2013). A noticeable difference in working with ordinal data the software will compute thresholds which are used to map the categorical variables onto an assumed underlying normal distribution of latent item responses and therefore create a set of latent correlations. This process is can be conceptualized as the reverse of the process used to create ordered categorical data from the original continuous data show in prior steps.

The concept of thresholds can be visualized by plotting the distribution of values for an item both in its continuous and categorical form. For this example, responses to I1 in the continuous data are visualized with a density plot (Figure ESI21a) and I1 responses in the categorical data are visualized with a bar plot (Figure ESI21b) using the code below.

```
plot(density(combined$I1),
      main="Density Plot for Combined Data Item I1 - Continuous",
      ylab="Frequency", xlab="Response")

barplot(prop.table(table(combined.ord$I1)),
         main="Frequency Plot for Combined Data Item I1 - Ordinal",
         ylab="Frequency", xlab="Response")
```

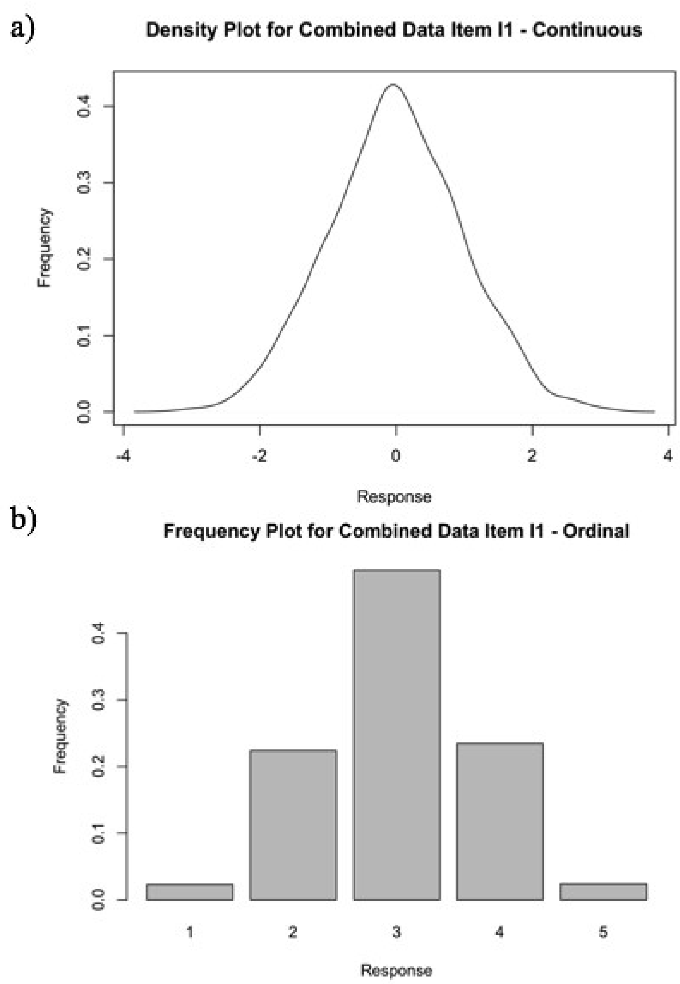

Figure ESI21. Density plot of continuous I1 responses (a) and frequency plot of categorical I1 responses (b)

Visual inspection of the two plots shows how the original continuous distribution aligns with the categorical data in that the middle responses have higher response frequencies and the extreme responses have lower response frequencies. When the ordinal data in Figure ESI21b are used to estimate a factor model, the software will assume the categorical data are representative of an underlying continuous variable (DiStefano and Morgan, 2014) and determine cut points, called thresholds, where the unobserved continuous distribution would have been divided to create the observed categorical distribution.

Since the categorical data used in this example were created from continuous data, we are able find the true cut points using the same code as before.

```
summary(cut(combined$I1, breaks=5))
```

Plotting these cut points (–1.97, –0.672, 0.624, and 1.92) on the continuous distribution (Figure ESI22) shows how the categorical data were simulated, and also provides insight into how the factor analysis itself will identify thresholds in the categorical data.

```
plot(density(combined$I1), main="Density Plot for Combined Data
      Item I1 - Continuous",ylab="Frequency", xlab="Response")
abline(v=c(-1.97, -0.672, 0.624, 1.92), col="grey")
```

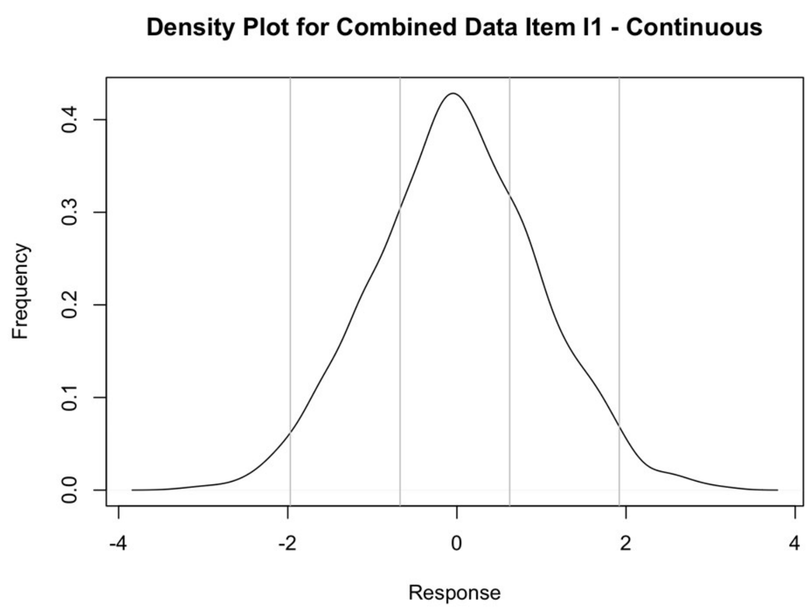

Figure ESI22. Density plot of continuous I1 responses showing cut points used to create categorical data.

Estimating Models with Ordered Categorical Data in R and Mplus

Running the factor models in R and also exporting the data for running in Mplus will provide an opportunity to see the threshold values established by the software. Full measurement invariance testing steps will be described in later sections. Both programs will automatically switch to the correct estimator (WLSMV) when informed that the data are not continuous. In lavaan syntax the argument `ordered` is used.

```
combined.ord.cfa<-cfa(data = combined.ord, model = model.test,
                      ordered=c("I1", "I2", "I3", "I4", "I5",
                      "I6", "I7", "I8", "I9", "I10", "I11",
                      "I12"))
summary(combined.ord.cfa, standardized=TRUE, fit.measures=TRUE)

combined.ord$group<-combined.ord$group %>% as.factor() %>%
                    as.numeric()
prepareMplusData(combined.ord, filename="CombinedOrdinal.dat",
                 inpfile = T, keepCols=c("I1", "I2", "I3",
                 "I4","I5", "I6", "I7", "I8", "I9", "I10", "I11",
                 "I12", "group"))
```

In Mplus the variables are specified as `categorical`.

```
TITLE: Combined Ordinal Data - CFA Model
DATA: FILE = "CombinedOrdinal.dat";
VARIABLE:
NAMES = I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12 group;
MISSING=.;

USEVARIABLES ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
CATEGORICAL ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;

MODEL:
IC BY I1 I2 I3 I4;
CC BY I5 I6 I7 I8;
AC BY I9 I10 I11 I12;

OUTPUT:
STANDARDIZED;
```

The full output of both programs can be examined to confirm similarities in how the data are treated as well as the matched fit indices and model parameters. Figure ESI23 shows the threshold values calculated by each program, indicated with the `t` notation in R and the `$` notation in Mplus. As expected, the thresholds for I1 are similar to those used to create the categorical data from the continuous, even though neither R or Mplus had access to the continuous data when generating the threshold values.

a)

```
Thresholds:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
    I1|t1            -1.919    0.058  -33.205    0.000   -1.919   -1.919
    I1|t2            -0.645    0.030  -21.314    0.000   -0.645   -0.645
    I1|t3             0.674    0.030   22.131    0.000    0.674    0.674
    I1|t4             1.927    0.058   33.130    0.000    1.927    1.927
```

b)

```
 Thresholds
    I1$1               -1.919      0.058    -33.213      0.000
    I1$2               -0.645      0.030    -21.319      0.000
    I1$3                0.674      0.030     22.137      0.000
    I1$4                1.927      0.058     33.138      0.000
```

Figure ESI23. Threshold values from R (a) and Mplus (b)

Data Model Fit for Ordered Categorical Data with WLSMV Estimator

The fit index cut off values recommended by Hu and Bentler (1999) were based on work using the maximum likelihood (ML) estimator which is appropriate for continuous data. Since a different estimator is used with categorical data, it is not appropriate to use the same Hu and Bentler recommendations for fit index cut off values. Simulation studies with the WLSMV estimator have indicated that more rigorous cut off values are best, particularly when the data contain a small number of categories or are severely nonnormal (Yu, 2002; Beauducel and Herzberg, 2006; DiStefano and Morgan, 2014). Recommendations for fit index values with the WLSMV estimator are CFI ≥ 0.95 and RMSEA ≤ 0.05. The SRMR is not recommended with the WLSMV estimator. In the context of invariance testing, less work has been done to determine recommended values for change in CFI and RMSEA values between models compared to the ML estimator. As with the fit indices themselves, simulation studies suggest either using more rigorous ΔCFI and ΔRMSEA values than those used with ML estimation or providing multiple sources of justification for acceptable data-model fit potentially using different estimators to see if similar conclusions about invariance would be drawn (Sass *et al.*, 2014).

Invariance Testing with R – Ordered Categorical Data

Measurement invariance testing in R with categorical data can be conducted following similar steps as those used for continuous data. However, it should be noted that other researchers have advocated for a different order of steps or different sets of constraints when working with categorical data (Millsap and Yun-Tein, 2004; Wu and Estabrook, 2016; Svetina *et al.*, 2019). The primary differences when working with categorical data compared to continuous are that the ordinal nature of the data must be specified in order for the correct estimator to be used, and thresholds must be constrained along with other model parameters during invariance testing steps.

Also, unique to working with categorical data, a decision must be made about scaling of the underlying latent normal distribution for each set of item responses using either delta or theta scaling. In delta scaling the total variance of the latent response is set to 1 and in theta scaling the variance of the residual term is set to 1. These decisions primarily influence how the model parameters are identified. Theta scaling is appropriate for invariance research (Millsap and Yun-Tein, 2004) and was chosen for the analysis here, but it is possible to convert parameters between delta and theta scaling (Finney and DiStefano, 2013). Since theta scaling affects the residual terms, Step 4 of invariance testing (strict) is not necessary with categorical data when following this method.

The steps taken in this ESI will parallel those used previously for continuous data. The data used in this section are the categorical version of the continuous data used in previous examples where the mean for I3 was changed in the STEM majors group. The code for all steps of invariance testing in R with categorical data are specified below and the fit statistics are summarized in Table ESI3 using the WLSMV output from lavaan as given in the `Robust` column. Fit statistics for models using the other categorical datasets are provided in Tables ESI4 & ESI5.

Step 0: Establishing Baseline Model

The baseline model for each group is specified in the same way as the continuous data but now using the ordinal data set and specifying which variables are ordered categorical as well as the use of the theta parameterization. The same three factor model used for the continuous data is used for the categorical data.

```
STEM.step0.ord<-cfa(data = combined.invar.mean.ord %>%
                     filter(group==STEM), model=model.test,
                     ordered=c("I1", "I2", "I3", "I4", "I5", "I6",
                     "I7", "I8", "I9", "I10", "I11", "I12"),
                     parameterization="theta")

summary(STEM.step0.ord, standardized=TRUE, fit.measures=TRUE)

nonSTEM.step0.ord<-cfa(data=combined.invar.mean.ord %>%
                        filter(group=="nonSTEM"),
                        model=model.test, ordered=c("I1", "I2",
                        "I3", "I4", "I5", "I6", "I7", "I8", "I9",
                        "I10", "I11", "I12"),
                        parameterization="theta")

summary(nonSTEM.step0.ord, standardized=TRUE, fit.measures=TRUE)
```

Step 1: Configural Invariance

Configural invariance uses data from both groups while specifying the grouping variable.

```
step1.comb.mean.ord<-cfa(data=combined.invar.mean.ord,
                          group="group", model=model.test,
                          ordered=c("I1", "I2", "I3", "I4", "I5",
                          "I6", "I7", "I8", "I9", "I10", "I11",
                          "I12"), parameterization="theta")

 summary(step1.comb.mean.ord, standardized=TRUE,
         fit.measures=TRUE)
```

Step 2: Metric Invariance (Weak)

Metric invariance is tested by holding the loadings equal across groups.

```
step2.comb.mean.ord<-cfa(data=combined.invar.mean.ord,
                          group="group", model=model.test,
                          ordered=c("I1", "I2", "I3", "I4", "I5",
                          "I6", "I7", "I8", "I9", "I10", "I11",
                          "I12"), group.equal=c("loadings"),
                          parameterization="theta")

summary(step2.comb.mean.ord, standardized=TRUE,
fit.measures=TRUE)
```

Step 3: Scalar Invariance (Strong)

Adding the constraint of equal thresholds across groups is similar to holding intercepts equal to test for scalar invariance in continuous data.

```
step3.comb.mean.ord<-cfa(data=combined.invar.mean.ord,
                         group="group", model=model.test,
                         ordered=c("I1", "I2", "I3", "I4", "I5",
                         "I6", "I7", "I8", "I9", "I10", "I11",
                         "I12"), group.equal=c("loadings",
                         "thresholds"), parameterization="theta")

summary(step3.comb.mean.ord, standardized=TRUE,
fit.measures=TRUE)
```

Table ESI3. Measurement Invariance Testing for the PRCQ Instrument Comparing STEM Majors and Non-STEM Majors With `combined.invar.mean` Simulated Categorical Data for Illustration

Step	**Testing level**	**χ2**	***df***	***p*-value**	**CFI**	**RMSEA**	**Δχ2**	**Δ*df***	***p*-value**	**ΔCFI**	**ΔRMSEA**
0	STEM majors Baseline	81	51	0.005	0.996	0.024	-	-	-	-	-
0	Non-STEM majors Baseline	61	51	0.162	0.999	0.014	-	-	-	-	-
1	Configural	142	102	0.006	0.997	0.020	-	-	-	-	-
2	Metric	145	111	0.017	0.998	0.018	3	9	0.231	0.001	0.002
3	Scalar	869	144	< 0.001	0.953	0.071	724	9	< 0.001	0.045	0.053

Note. STEM majors n = 1000. Non-STEM majors n = 1000. Simulated data was used and altered at the scalar level (intercepts) for illustrative purposes; fit indices are from R.

Invariance Testing with Mplus – Ordered Categorical Data

Following the previously shown steps, the categorical data in R are exported to Mplus by first converting the group variable from a text format into a numeric format.

```
combined.invar.mean.ord$group<-combined.ord$group %>% as.factor()
                                %>% as.numeric()

prepareMplusData(combined.invar.mean.ord,
                 filename="CombinedInvarMeanOrdinal.dat",
                 inpfile = T, keepCols=c("I1", "I2", "I3",
                 "I4","I5", "I6", "I7", "I8", "I9", "I10",
                 "I11", "I12", "group"))
```

As with lavaan, the default estimator in Mplus is ML but the software will adjust to an appropriate estimator for ordinal data (WLSMV) by specifying the item variables as categorical. The call for theta parameterization is also added and the models are specified separately for each group. Following these steps for R and Mplus should provide similar fit indices and model parameters.

Step 0: Establishing Baseline Model

```
TITLE: Categorical STEM Majors Group Step 0
DATA: FILE = "CombinedInvarMeanOrdinal.dat";
VARIABLE:
NAMES = I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12 group;
MISSING=.;

USEVARIABLES ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
CATEGORICAL ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
USEOBSERVATIONS are group==2;

ANALYSIS: PARAMETERIZATION=THETA;

MODEL:
IC BY I1 I2 I3 I4;
CC BY I5 I6 I7 I8;
AC BY I9 I10 I11 I12;

OUTPUT:
STANDARDIZED;


TITLE: Categorical Non-STEM Majors Group Step 0
DATA: FILE = "CombinedInvarMeanOrdinal.dat";
VARIABLE:
NAMES = I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12 group;
MISSING=.;

USEVARIABLES ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
CATEGORICAL ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
USEOBSERVATIONS are group==1;

ANALYSIS: PARAMETERIZATION=THETA;

MODEL:
IC BY I1 I2 I3 I4;
CC BY I5 I6 I7 I8;
AC BY I9 I10 I11 I12;

OUTPUT:
STANDARDIZED;
```

Step 1: Configural Invariance

By default, Mplus will constrain thresholds equal across groups so this must be released by freeing all thresholds for all variables. The notation to free the thresholds uses the `$` character. Four thresholds must be freed since four thresholds would be required to divide the underlying continuous distribution into five categories. As was done with the continuous data, the factor means are set to zero. The error variances are set to one for categorical data, in line with theta parameterization.

```
TITLE: Categorical Combined Dataset with Mean Differences Step 1
(Configural)
DATA: FILE = "CombinedInvarMeanOrdinal.dat";
VARIABLE:
NAMES = I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12 group;
CATEGORICAL ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
USEVARIABLES ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
GROUPING = group (1 = NonSTEM 2 = STEM);

ANALYSIS: PARAMETERIZATION=THETA;

MODEL:
! Model with standardized loading of first item on each factor
  IC BY I1@1 I2 I3 I4;
  CC BY I5@1 I6 I7 I8;
  AC BY I9@1 I10 I11 I12;

! Freeing Thresholds
  [I1$1-I12$1*];
  [I1$2-I12$2*];
  [I1$3-I12$3*];
  [I1$4-I12$4*];

! Set factor means to 0
  [IC@0];
  [CC@0];
  [AC@0];

! Set error variances to 1
  I1-I12@1

MODEL STEM:
  IC BY I1@1 I2 I3 I4;
  CC BY I5@1 I6 I7 I8;
  AC BY I9@1 I10 I11 I12;

! Freeing Thresholds
  [I1$1-I12$1*];
  [I1$2-I12$2*];
  [I1$3-I12$3*];
  [I1$4-I12$4*];
```

```
! Set factor means to 0
  [IC@0];
  [CC@0];
  [AC@0];

! Set error variances to 1
  I1-I12@1

OUTPUT:
STANDARDIZED;
```

Step 2: Metric Invariance (Weak)

Loadings are constrained equal across groups by assigning the same name to the parameters in both groups. This is the same method used for invariance testing with the continuous data.

```
TITLE: Categorical Combined Dataset with Mean Differences Step 2
DATA: FILE = "CombinedInvarMeanOrdinal.dat";
VARIABLE:
NAMES = I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12 group;
CATEGORICAL ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
USEVARIABLES ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
GROUPING = group (1 = NonSTEM 2 = STEM);

ANALYSIS: PARAMETERIZATION=THETA;

MODEL:
! Model with standardized loading of first item on each factor
! Assigning a parameter name to each loading value (L1-L12)
  IC BY I1@1 I2 I3 I4 (L1-L4);
  CC BY I5@1 I6 I7 I8 (L5-L8);
  AC BY I9@1 I10 I11 I12 (L9-L12);

! Freeing Thresholds
  [I1$1-I12$1*];
  [I1$2-I12$2*];
  [I1$3-I12$3*];
  [I1$4-I12$4*];

! Set factor means to 0
  [IC@0];
  [CC@0];
  [AC@0];

! Set error variances to 1
  I1-I12@1

MODEL STEM:
  IC BY I1@1 I2 I3 I4 (L1-L4);
  CC BY I5@1 I6 I7 I8 (L5-L8);
  AC BY I9@1 I10 I11 I12 (L9-L12);
```

```
  ! Freeing Thresholds
  [I1$1-I12$1*];
  [I1$2-I12$2*];
  [I1$3-I12$3*];
  [I1$4-I12$4*];

  ! Set factor means to 0
  [IC@0];
  [CC@0];
  [AC@0];

 ! Set error variances to 1
  I1-I12@1

OUTPUT:
STANDARDIZED;
```

Step 3: Scalar Invariance (Strong)

Mplus and lavaan differ in their default settings when thresholds are constrained equal across groups. To mimic the lavaan output the factor means and error variance terms for the second group are freed in the Mplus code. Freeing these parameters also aligns scalar invariance testing in the categorical data with the same step for the continuous data. Recall that the goal of Step 3 is to determine if the factors are being measured on the same scale in each group so that factor means can be compared across groups. Therefore, one group should have a mean of zero in order to function as a reference while the mean of the other group is freely estimated.

```
TITLE: Categorical Combined Dataset with Mean Differences Step 3
DATA: FILE = "CombinedInvarMeanOrdinal.dat";
VARIABLE:
NAMES = I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12 group;
CATEGORICAL ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
USEVARIABLES ARE I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11 I12;
GROUPING = group (1 = NonSTEM 2 = STEM);

ANALYSIS: PARAMETERIZATION=THETA;

MODEL:
  IC BY I1@1 I2 I3 I4 (L1-L4);
  CC BY I5@1 I6 I7 I8 (L5-L8);
  AC BY I9@1 I10 I11 I12 (L9-L12);

  [I1$1-I12$1*];
  [I1$2-I12$2*];
  [I1$3-I12$3*];
  [I1$4-I12$4*];

  [IC@0];
  [CC@0];
  [AC@0];

  I1-I12@1
```

```
MODEL STEM:
  IC BY I1@1 I2 I3 I4 (L1-L4);
  CC BY I5@1 I6 I7 I8 (L5-L8);
  AC BY I9@1 I10 I11 I12 (L9-L12);

! Fix thresholds equal by not specifying for this group

! Set factor means free
[IC*];
[CC*];
[AC*];

! Set error variances free
I1-I12*

OUTPUT:
STANDARDIZED;
```

Fit Indices for Invariance Testing Steps with other Simulated Categorical Data

Tables ESI4 & 5 show the data-model fit output from R produced from following the previous steps with the two other categorical datasets: `combined.ord` and `combined.invar.load.ord`.

Table ESI4. Measurement Invariance Testing for the PRCQ Instrument Comparing STEM Majors and Non-STEM Majors With `combined.ord` Simulated Categorical Data for Illustration

Step	**Testing level**	**$\chi2$**	***df***	***p*-value**	**CFI**	**RMSEA**	**$\Delta\chi2$**	**Δdf**	***p*-value**	**ΔCFI**	**ΔRMSEA**
0	STEM majors Baseline	81	51	0.005	0.996	0.024	-	-	-	-	-
0	Non-STEM majors Baseline	61	51	0.162	0.999	0.014	-	-	-	-	-
1	Configural	142	102	0.006	0.997	0.020	-	-	-	-	-
2	Metric	145	111	0.017	0.998	0.018	3	9	0.964	0.001	0.002
3	Scalar	869	144	<0.001	0.953	0.071	724	33	< 0.001	0.045	0.053

Note. STEM majors *n* = 1000. Non-STEM majors *n* = 1000. Simulated data was used and altered at the scalar level (intercepts) for illustrative purposes; fit indices are from R.

Table ESI5. Measurement Invariance Testing for the PRCQ Instrument Comparing STEM Majors and Non-STEM Majors With `combined.invar.load.ord` Simulated Categorical Data for Illustration

Step	Testing level	$\chi 2$	*df*	*p*-value	CFI	RMSEA	$\Delta\chi 2$	Δdf	*p*-value	ΔCFI	ΔRMSEA
0	STEM majors Baseline	81	51	0.005	0.996	0.024	-	-	-	-	-
0	Non-STEM majors Baseline	40	51	0.869	1.000	0.000	-	-	-	-	-
1	Configural	119	102	0.120	0.999	0.013	-	-	-	-	-
2	Metric	383	111	< 0.001	0.982	0.050	264	9	< 0.001	0.017	0.037
3	Scalar	1305	144	< 0.001	0.925	0.090	922	33	< 0.001	0.057	0.040

Note. STEM majors $n = 1000$. Non-STEM majors $n = 1000$. Simulated data was used and altered at the scalar level (intercepts) for illustrative purposes; fit indices are from R.

ESI References

Beauducel A. and Herzberg P. Y., (2006), On the Performance of Maximum Likelihood Versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA. *Struct. Equ. Model. A Multidiscip. J.*, **13**(2), 186–203.

Bontempo D. E. and Hofer S. M., (2007), Assessing Factorial Invariance in Cross-Sectional and Longitudinal Studies., in Ong A. D. and van Dulmen M. H. M. (eds.), *Series in positive psychology. oxford handbook of methods in positive psychology*. Oxford University Press, pp. 153–175.

Byrne B. M., (2004), Testing for multigroup invariance using AMOS Graphics: A road less traveled. *Struct. Equ. Model. A Multidiscip. J.*, **11**(2), 272–300.

DiStefano C. and Morgan G. B., (2014), A Comparison of Diagonal Weighted Least Squares Robust Estimation Techniques for Ordinal Data. *Struct. Equ. Model.*, **21**, 425–438.

Finney S. J. and DiStefano C., (2013), Non-normal and categorical data in structural equation modeling., in Hancock G. R. and Mueller R. O. (eds.), *Structural equation modeling: a second course*. Charlotte, NC: Information Age Publishing, pp. 439–492.

Hallquist M. N. and Wiley J. F., (2018), MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus. *Struct. Equ. Model.*, **25**(4), 621–638.

Hancock G. R., Stapleton L. M., and Arnold-Berkovits I., (2009), The tenuousness of invariance tests within multisample covariance and mean structure models., in *Structural equation modeling in educational research: concepts and applications*., pp. 137–174.

Hirschfeld G. and Von Brachel R., (2014), Multiple-Group confirmatory factor analysis in R – A tutorial in measurement invariance with continuous and ordinal. *Pract. Assessment, Res. Eval.*, **19**(7), 1–11.

Hu L. and Bentler P. M., (1999), Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model. A Multidiscip. J.*, **6**(1), 1–55.

Komperda R., (2017), Likert-Type Survey Data Analysis with R and RStudio., in Gupta T. (ed.), *Computer-aided data analysis in chemical education research (cadacer): advances and avenues*. Washington, DC: ACS Symposium Series; American Chemical Society, pp. 91–116.

Millsap R. E. and Yun-Tein J., (2004), Assessing Factorial Invariance in Ordered-Categorical Measures. *Multivariate Behav. Res.*, **39**(3), 479–515.

Muthén L. K. and Muthén B. O., (2017), Mplus User's Guide, Eighth. Los Angeles, CA: Muthén & Muthén.

Narayanan A., (2012), A review of eight software packages for structural equation modeling. *Am. Stat.*, **66**(2), 129–138.

R Core Team, (2019), *R: A language and environment for statistical computing*, [Computer software].

Revelle W., (2018), *psych: procedures for psychological, psychometric, and personality research*, [Computer software].

Rosseel Y., (2020), The lavaan Project. http://lavaan.ugent.be/

Rosseel Y., (2012), lavaan: An R Package for Structural Equation Modeling. *J. Stat. Softw.*, **48**(2), 1–36.

Sass D. A., Schmitt T. A., and Marsh H. W., (2014), Evaluating Model Fit With Ordered Categorical Data Within a Measurement Invariance Framework: A Comparison of Estimators. *Struct. Equ. Model.*, **21**(2), 167–180.

Schneider W. J., (2019), *simstandard: generate standardized data*, [Computer software].

Svetina D., Rutkowski L., and Rutkowski D., (2019), Multiple-Group Invariance with Categorical Outcomes Using Updated Guidelines: An Illustration Using Mplus and the lavaan/semTools Packages. *Struct. Equ. Model.*, **0**(0), 1–20.

Wei T. and Simko V., (2017), *R package "corrplot": visualization of a correlation matrix,* [Computer software].

Wickham H., (2007), Reshaping Data with the reshape Package. *J. Stat. Softw.*, **21**(12), 1–20.

Wickham H., (2016), ggplot2: Elegant Graphics for Data Analysis, New York: Springer-Verlag.

Wickham H., François R., Henry L., and Müller K., (2019), *dplyr: a grammar of data manipulation,* [Computer software].

Wu H. and Estabrook R., (2016), Identification of Confirmatory Factor Analysis Models of Different Levels of Invariance for Ordered Categorical Outcomes. *Psychometrika*, **81**(4), 1014–1045.

Yu C.-Y., (2002), Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes.