Example 4.4 - Invariance of the Wechsler Intelligence Scale

```
library("lavaan")
```

This example comes from Beaujean, Freeman, Youngstrom, and Carlson (2012). They examined if the structure of the Wechsler intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) was the same in children with and without manic symptoms. They obtained sample covariance matrices of the WISC from two samples: a sample of manic depressive patients (N = 81) and a norm group (N = 200). We are going to assess whether the WISC subscales are measurement invariance across these two groups.

```
manic.cov <- lav_matrix_lower2full(c(
  9.364,
  7.777, 12.461,
  6.422,  8.756, 10.112,
  5.669,  7.445,  6.797, 8.123,
  3.048,  4.922,  4.513, 4.116, 6.200,
  3.505,  4.880,  4.899, 5.178, 5.114, 15.603,
  3.690,  5.440,  5.220, 3.151, 3.587,  6.219, 11.223,
  3.640,  4.641,  4.877, 3.568, 3.819,  5.811,  6.501, 9.797
))
manic.means <- c(10.09, 12.07, 10.25, 9.96, 10.90, 11.24, 10.30, 10.44)

norming.cov <- lav_matrix_lower2full(c(
  9.610,
  5.844, 8.410,
  6.324, 6.264, 9.000,
  4.405, 4.457, 5.046, 8.410,
  4.464, 4.547, 4.512, 3.712, 10.240,
  3.478, 2.967, 2.970, 2.871,  3.802, 10.890,
  5.270, 4.930, 4.080, 3.254,  5.222,  3.590, 11.560,
  4.297, 4.594, 4.356, 3.158,  4.963,  3.594,  6.620, 10.890
))
norming.means <- c(10.10, 10.30, 9.80, 10.10, 10.10, 10.10, 9.90, 10.20)

wisc3.names <- c("Info", "Sim", "Vocab","Comp", "PicComp", "PicArr",
                 "BlkDsgn", "ObjAsmb")
colnames(norming.cov) <- rownames(norming.cov)  <- colnames(manic.cov) <-
  rownames(manic.cov) <- names(norming.means) <- names(manic.means) <-
  wisc3.names

wisc3.model <- '
  VC =~ Info + Sim + Vocab + Comp
  VS =~ PicComp + PicArr + BlkDsgn + ObjAsmb
  VC ~ NA*1
  VS ~ NA*1
  Info ~ 0*1
  PicComp ~ 0*1
'
```

Note that I set the latent factor means to be freely estimated, and set the intercepts of the first indicators to 0. The default in **lavaan** is to use the marker variable approach for identifying the covariance structure, but to use the standardized LV approach for identifying the mean structure. I prefer (and most researchers in the field would agree) to use the same identification approach for both structures.

I prefer the marker variable approach, because the two groups may likely differ in terms of their average levels and the extent of inter-personal variation in intelligence.

Configural invariance

```r
manic.fit <- cfa(wisc3.model, sample.cov = manic.cov, sample.nobs = 81,
                 sample.mean = manic.means, meanstructure = TRUE)
summary(manic.fit, standardized = TRUE)
```

```
## lavaan 0.6-6 ended normally after 106 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                         25
##
##   Number of observations                            81
##
## Model Test User Model:
##
##   Test statistic                                29.169
##   Degrees of freedom                                19
##   P-value (Chi-square)                           0.063
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   VC =~
##     Info              1.000                               2.347    0.772
##     Sim               1.330    0.153    8.687    0.000    3.121    0.890
##     Vocab             1.189    0.138    8.613    0.000    2.791    0.883
##     Comp              1.015    0.125    8.129    0.000    2.382    0.841
##   VS =~
##     PicComp           1.000                               1.788    0.723
##     PicArr            1.437    0.274    5.246    0.000    2.570    0.655
##     BlkDsgn           1.322    0.234    5.641    0.000    2.364    0.710
##     ObjAsmb           1.285    0.220    5.830    0.000    2.297    0.738
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   VC ~~
##     VS                3.086    0.772    3.997    0.000    0.735    0.735
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
```

```
##     VC                10.090    0.338   29.861    0.000    4.300    4.300
##     VS                10.900    0.275   39.643    0.000    6.096    6.096
##    .Info               0.000                               0.000    0.000
##    .PicComp            0.000                               0.000    0.000
##    .Sim               -1.352    1.581   -0.855    0.393   -1.352   -0.385
##    .Vocab             -1.749    1.426   -1.226    0.220   -1.749   -0.553
##    .Comp              -0.283    1.290   -0.219    0.826   -0.283   -0.100
##    .PicArr            -4.425    3.016   -1.467    0.142   -4.425   -1.127
##    .BlkDsgn           -4.109    2.580   -1.593    0.111   -4.109   -1.234
##    .ObjAsmb           -3.563    2.426   -1.469    0.142   -3.563   -1.145
##
## Variances:
##                     Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .Info               3.742    0.673    5.564    0.000    3.742    0.405
##    .Sim                2.564    0.605    4.237    0.000    2.564    0.208
##    .Vocab              2.200    0.502    4.379    0.000    2.200    0.220
##    .Comp               2.348    0.467    5.027    0.000    2.348    0.293
##    .PicComp            2.926    0.592    4.945    0.000    2.926    0.478
##    .PicArr             8.806    1.631    5.398    0.000    8.806    0.571
##    .BlkDsgn            5.497    1.089    5.046    0.000    5.497    0.496
##    .ObjAsmb            4.399    0.916    4.804    0.000    4.399    0.455
##     VC                 5.507    1.369    4.024    0.000    1.000    1.000
##     VS                 3.198    0.924    3.462    0.001    1.000    1.000
```

```r
fit.indices <- c("chisq", "df", "pvalue", "cfi", "rmsea", "srmr", "aic",
                 "bic")
fitMeasures(manic.fit, fit.indices)
```

```
##    chisq       df   pvalue      cfi    rmsea     srmr      aic      bic
##   29.169   19.000    0.063    0.971    0.081    0.047 3019.261 3079.122
```

All loadings are significant and substantial. The χ^2, CFI, and SRMR indicate good fit in the manic group. The RMSEA indicates less than adequate fit. All loadings are substantial and significant. The two factors also correlate substantially and significantly.

```r
norming.fit <- cfa(wisc3.model, sample.cov = norming.cov, sample.nobs = 200,
                   sample.mean = norming.means, meanstructure = TRUE)
parameterestimates(norming.fit, standardized = TRUE)[ , c(1:5, 7, 11)]
```

```
##        lhs op     rhs    est    se pvalue std.all
## 1       VC =~    Info  1.000 0.000     NA   0.789
## 2       VC =~     Sim  0.997 0.079  0.000   0.841
## 3       VC =~   Vocab  1.045 0.082  0.000   0.852
## 4       VC =~    Comp  0.768 0.083  0.000   0.648
## 5       VS =~ PicComp  1.000 0.000     NA   0.684
## 6       VS =~  PicArr  0.715 0.122  0.000   0.474
## 7       VS =~ BlkDsgn  1.149 0.135  0.000   0.739
## 8       VS =~ ObjAsmb  1.100 0.130  0.000   0.730
## 9       VC ~1         10.100 0.219  0.000   4.139
## 10      VS ~1         10.100 0.226  0.000   4.628
## 11    Info ~1          0.000 0.000     NA   0.000
## 12 PicComp ~1          0.000 0.000     NA   0.000
## 13    Info ~~    Info  3.609 0.455  0.000   0.377
## 14     Sim ~~     Sim  2.450 0.354  0.000   0.293
## 15   Vocab ~~   Vocab  2.453 0.370  0.000   0.274
## 16    Comp ~~    Comp  4.857 0.533  0.000   0.580
```

```
## 17 PicComp ~~ PicComp  5.426 0.675  0.000   0.533
## 18  PicArr ~~  PicArr  8.398 0.897  0.000   0.775
## 19 BlkDsgn ~~ BlkDsgn  5.215 0.716  0.000   0.453
## 20 ObjAsmb ~~ ObjAsmb  5.069 0.682  0.000   0.468
## 21      VC ~~      VC  5.953 0.928  0.000   1.000
## 22      VS ~~      VS  4.763 0.952  0.000   1.000
## 23      VC ~~      VS  4.103 0.661  0.000   0.771
## 24     Sim ~1          0.230 0.817  0.779   0.079
## 25   Vocab ~1         -0.755 0.845  0.372  -0.252
## 26    Comp ~1          2.344 0.855  0.006   0.810
## 27  PicArr ~1          2.875 1.257  0.022   0.874
## 28 BlkDsgn ~1         -1.704 1.381  0.217  -0.502
## 29 ObjAsmb ~1         -0.913 1.335  0.494  -0.277
```

```
fitMeasures(norming.fit, fit.indices)
```

```
##    chisq       df    pvalue      cfi     rmsea     srmr      aic      bic
##   24.211   19.000    0.188    0.992    0.037    0.029 7564.068 7646.526
```

We see better model fit in the norm group. So perhaps we should not just assume configural invariance here. In such a case, I would first inspect the fitted model, residuals and modification indices in each of the two groups.

```
residuals(manic.fit, type = "cor")
```

```
## $type
## [1] "cor.bollen"
##
## $cov
##         Info   Sim    Vocab  Comp   PicCmp PicArr BlkDsg ObjAsm
## Info     0.000
## Sim      0.033  0.000
## Vocab   -0.021 -0.006  0.000
## Comp     0.001 -0.008  0.007  0.000
## PicComp -0.010  0.087  0.101  0.133  0.000
## PicArr  -0.081 -0.078 -0.035  0.055  0.047  0.000
## BlkDsgn -0.043 -0.005  0.029 -0.109 -0.083  0.005  0.000
## ObjAsmb -0.039 -0.063  0.010 -0.057 -0.044 -0.013  0.096  0.000
##
## $mean
##    Info     Sim   Vocab    Comp PicComp  PicArr BlkDsgn ObjAsmb
##       0       0       0       0       0       0       0       0
```

```
modindices(manic.fit, sort = TRUE)[1:6, ]
```

```
##        lhs op     rhs    mi    epc sepc.lv sepc.all sepc.nox
## 65 BlkDsgn ~~ ObjAsmb 7.391  2.227   2.227    0.453    0.453
## 58    Comp ~~ BlkDsgn 7.116 -1.304  -1.304   -0.363   -0.363
## 30      VC =~ PicComp 6.621  0.484   1.137    0.459    0.459
## 61 PicComp ~~ BlkDsgn 4.992 -1.448  -1.448   -0.361   -0.361
## 57    Comp ~~  PicArr 4.737  1.310   1.310    0.288    0.288
## 56    Comp ~~ PicComp 3.153  0.639   0.639    0.244    0.244
```

In the manic group, there are standardized residuals $> .1$ for `PicComp ~~ Comp`, `Piccomp ~~ Vocab`, `Comp ~~ BlkDsgn`. Modification indices suggest adding `BlkDsgn ~~ ObjAsmb`.

```r
residuals(norming.fit, type = "cor")
```

```
## $type
## [1] "cor.bollen"
##
## $cov
##         Info   Sim    Vocab  Comp   PicCmp PicArr BlkDsg ObjAsm
## Info     0.000
## Sim     -0.014  0.000
## Vocab    0.008  0.003  0.000
## Comp    -0.021 -0.015  0.028  0.000
## PicComp  0.034  0.047  0.021  0.059  0.000
## PicArr   0.052  0.003 -0.011  0.063  0.036  0.000
## BlkDsgn  0.051  0.021 -0.085 -0.039 -0.026 -0.031  0.000
## ObjAsmb -0.023  0.007 -0.039 -0.034 -0.029 -0.016  0.051  0.000
##
## $mean
##    Info     Sim   Vocab    Comp PicComp  PicArr BlkDsgn ObjAsmb
##       0       0       0       0       0       0       0       0
```

```r
modindices(norming.fit, sort = TRUE)[1:6, ]
```

```
##        lhs op     rhs    mi    epc sepc.lv sepc.all sepc.nox
## 54   Vocab ~~ BlkDsgn 7.962 -0.988  -0.988   -0.276   -0.276
## 65 BlkDsgn ~~ ObjAsmb 7.148  1.670   1.670    0.325    0.325
## 36      VS =~   Vocab 5.765 -0.359  -0.784   -0.262   -0.262
## 43    Info ~~ BlkDsgn 4.150  0.794   0.794    0.183    0.183
## 30      VC =~ PicComp 3.195  0.326   0.796    0.249    0.249
## 51   Vocab ~~    Comp 1.936  0.475   0.475    0.137    0.137
```

In the norming group, there are no standardized residuals $> .1$. Modification indices for the norming group also suggest adding `BlkDsgn ~~ ObjAsmb` (like in the manic group). These may indeed have something in common that is not shared by the other indictors of the Visuo-Spatial factor.

Should we add a correlated error to the model? That is a decision you, the researcher, have to make. Adding correlated errors violates the assumption of conditional independence: that conditional on the common factor(s), the observed indicators are independent. So I would rather not add a correlated error. Later on, I will perform a sensitivity analysis to check whether the addition of a correlated error would yield different conclusions.

For now, I conclude configural invariance is tenable and carry on with the multigroup analysis.

We first have to combine the covariance matrices, sample sizes and means of both groups into lists:

```r
combined.cov <- list(manic = manic.cov, norming = norming.cov)
combined.n <- list(manic = 81, norming = 200)
combined.means <- list(manic = manic.means, norming = norming.means)
```

Note that in practice, you will often analyse the whole dataset, so combining the means and covariances into lists is not necessary. With the raw data, you would specify the `data` and `group` arguments, instead of the `sample.cov`, `sample.nobs` and `sample.mean` arguments.

```r
configural.fit <- cfa(wisc3.model, sample.cov = combined.cov,
                      sample.nobs = combined.n, sample.mean = combined.means,
                      meanstructure = TRUE)
fitMeasures(configural.fit, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
```

```
##     53.380     38.000      0.050      0.985      0.054      0.034 10583.329 10765.247
```

These fit indices indicate configural invariance seems tenable, overall: RMSEA, CFI and SRMR indicate good fit, the p-value of the χ^2 is quite high. Next, I constrain the loadings to be equal across groups:

Metric invariance

```r
metric.fit <- cfa(wisc3.model, sample.cov = combined.cov,
                  sample.nobs = combined.n, sample.mean = combined.means,
                  meanstructure = TRUE, group.equal = "loadings")
fitMeasures(metric.fit, fit.indices)
```

```
##     chisq         df     pvalue        cfi      rmsea       srmr        aic        bic
##    65.992     44.000      0.018      0.979      0.060      0.055 10583.942 10744.029
```

```r
lavTestLRT(metric.fit, configural.fit)
```

```
## Chi-Squared Difference Test
##
##                Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## configural.fit 38 10583 10765 53.380
## metric.fit     44 10584 10744 65.992     12.613       6    0.04961 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

CFI has decreased, but not significantly (i.e., $\Delta CFI < .01$). RMSEA increased, but still indicates good fit. AIC shows a very slight increase, while BIC has decreased. The $\Delta\chi^2$ test indicates a significant difference in model fit. Note that the Δdf equals six: all (8) loadings were restricted to equality, but because we used the marker-variable approach to identify the scale of the LVs, two of the 8 loadings were already equal. Thus, 6 parameters were restricted to be equal in this step, yielding $\Delta df = 6$.

We can use modification indices to find out which parameter restriction causes the misfit. The `modificationindices()` function does not give modification indices for parameters that are restricted to equality anymore. You have to use the `lavTestScore()` function for that:

```r
lavTestScore(metric.fit)$uni
```

```
##
## univariate score tests:
##
##    lhs op   rhs    X2 df p.value
## 1 .p2. == .p31. 2.501  1   0.114
## 2 .p3. == .p32. 0.557  1   0.455
## 3 .p4. == .p33. 0.874  1   0.350
## 4 .p6. == .p35. 6.144  1   0.013
## 5 .p7. == .p36. 0.042  1   0.837
## 6 .p8. == .p37. 0.001  1   0.971
```

```r
pars <- parameterestimates(metric.fit)
pars[pars$label == ".p6.", 1:3]
```

```
##    lhs op    rhs
## 6   VS =~ PicArr
## 35  VS =~ PicArr
```

The modification indices suggest lifting the restriction on parameter 6, which is the factor loading of Picture Arrangement on the VS factor. We specify this parameter with the `group.partial` argument, to release the equality restriction imposed by the `group.equal` command:

```
metric.fit2 <- cfa(wisc3.model, sample.cov = combined.cov,
                   sample.nobs = combined.n, sample.mean = combined.means,
                   meanstructure = TRUE, group.equal = "loadings",
                   group.partial = "VS =~ PicArr")
fitmeasures(metric.fit2, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##    59.500    43.000     0.048     0.984     0.052     0.045 10579.449 10743.175
```

```
lavTestLRT(metric.fit2, configural.fit)
```

```
## Chi-Squared Difference Test
##
##                Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## configural.fit 38 10583 10765 53.380
## metric.fit2    43 10579 10743 59.499       6.12       5     0.2947
```

We obtained an adequately fitting partial metric invariance model. AIC and BIC are lower than in the configural invariance model; CFI and RMSEA show only slight increases compared to the configural invariant model. The $\Delta\chi^2$ test is no longer significant. Note that by lifting one equality restriction, the Δdf value increased by 1.

Scalar invariance

```
scalar.fit <- cfa(wisc3.model, sample.cov = combined.cov,
                  sample.nobs = combined.n, sample.mean = combined.means,
                  meanstructure = TRUE,
                  group.equal = c("loadings", "intercepts"),
                  group.partial = "VS =~ PicArr")
fitMeasures(scalar.fit, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##   104.570    49.000     0.000     0.946     0.090     0.060 10612.519 10754.415
```

```
lavTestLRT(metric.fit2, scalar.fit)
```

```
## Chi-Squared Difference Test
##
##             Df   AIC   BIC   Chisq Chisq diff Df diff Pr(>Chisq)
## metric.fit2 43 10579 10743  59.499
## scalar.fit  49 10612 10754 104.570     45.071       6  4.531e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to all fit indices but the SRMR, the fit of the scalar invariance model is not acceptable. Also, ΔCFI was $> .01$ and $\Delta\chi^2$ was significant, compared to the partial metric invariance model.

```
lavTestScore(scalar.fit)$uni
```

```
##
## univariate score tests:
##
##      lhs op   rhs     X2 df p.value
## 1   .p2. == .p31. 30.824  1   0.000
## 2   .p3. == .p32.  0.993  1   0.319
```

```
## 3   .p4. == .p33.  5.995  1   0.014
## 4   .p7. == .p36.  0.221  1   0.638
## 5   .p8. == .p37.  1.200  1   0.273
## 6  .p24. == .p53. 29.453  1   0.000
## 7  .p25. == .p54.  0.848  1   0.357
## 8  .p26. == .p55.  7.455  1   0.006
## 9  .p27. == .p56.  4.791  1   0.029
## 10 .p28. == .p57.  0.206  1   0.650
## 11 .p29. == .p58.  1.244  1   0.265
```

The equality restriction on parameters p2 and p24 seems to cause most misfit. Let's see which parameters they are:

```r
pars <- parameterestimates(scalar.fit)
pars[pars$label %in% c(".p2.", ".p24."), 1:3]
```

```
##    lhs op rhs
## 2   VC =~ Sim
## 24 Sim ~1
## 31  VC =~ Sim
## 53 Sim ~1
```

The most violating equality restriction seems to be on the loading of Similarities on the VC factor. But this loading did not cause any problems in the last step. Furthermore, the intercept of the Similarities subtest also seems to be quite problematic. Lifting that equality restriction may relieve the problem with the loading:

```r
scalar.fit2 <- cfa(wisc3.model, sample.cov = combined.cov,
                   sample.nobs = combined.n, sample.mean = combined.means,
                   group.equal = c("loadings", "intercepts"),
                   group.partial = c("VS =~ PicArr", "Sim~1"),
                   meanstructure = TRUE)
fitMeasures(scalar.fit2, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##    71.412    48.000     0.016     0.977     0.059     0.054 10581.361 10726.895
```

```r
lavTestLRT(metric.fit2, scalar.fit2)
```

```
## Chi-Squared Difference Test
##
##             Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## metric.fit2 43 10579 10743 59.499
## scalar.fit2 48 10581 10727 71.412     11.912       5    0.03601 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
lavTestScore(scalar.fit2)$uni
```

```
##
## univariate score tests:
##
##      lhs op   rhs    X2 df p.value
## 1   .p2. == .p31. 2.550  1   0.110
## 2   .p3. == .p32. 2.725  1   0.099
## 3   .p4. == .p33. 1.323  1   0.250
## 4   .p7. == .p36. 0.224  1   0.636
## 5   .p8. == .p37. 1.200  1   0.273
## 6  .p25. == .p54. 3.609  1   0.057
```

```
## 7  .p26. == .p55. 1.967  1   0.161
## 8  .p27. == .p56. 4.786  1   0.029
## 9  .p28. == .p57. 0.209  1   0.647
## 10 .p29. == .p58. 1.240  1   0.265
```

```
pars <- parameterestimates(scalar.fit)
pars[pars$label == ".p27.", 1:3]
```

```
##       lhs op rhs
## 27 PicArr ~1
## 56 PicArr ~1
```

Fit improved. The χ^2 value is still significant. The most violating parameter now appears to be the intercept of Picture Arrangement, so the problem with the loading of Similarities was solved.

```
scalar.fit3 <- cfa(wisc3.model, sample.cov = combined.cov,
                   sample.nobs = combined.n, sample.mean = combined.means,
                   group.equal = c("loadings", "intercepts"),
                   group.partial = c("VS =~ PicArr", "Sim~1", "PicArr~1"),
                   meanstructure = TRUE)
fitMeasures(scalar.fit3, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##    66.381    47.000     0.033     0.981     0.054     0.047 10578.330 10727.503
```

```
lavTestLRT(metric.fit2, scalar.fit3)
```

```
## Chi-Squared Difference Test
##
##             Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## metric.fit2 43 10579 10743 59.499
## scalar.fit3 47 10578 10728 66.381     6.8816       4     0.1423
```

Now, the model fits well and the difference with the partial metric invariance model is not significant anymore. Thus, partial scalar invariance is tenable. We now proceed with testing across-group equality of residual variances:

Uniqueness invariance

```
uniqueness.fit <- cfa(wisc3.model, sample.cov = combined.cov,
                      sample.nobs = combined.n,
                      sample.mean = combined.means,
                      group.equal=c("loadings", "intercepts", "residuals"),
                      group.partial = c("VS =~ PicArr", "Sim~1", "PicArr~1"))
fitMeasures(uniqueness.fit, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##    84.199    55.000     0.007     0.972     0.061     0.055 10580.148 10700.214
```

```
lavTestLRT(uniqueness.fit, scalar.fit3)
```

```
## Chi-Squared Difference Test
##
##                Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## scalar.fit3    47 10578 10728 66.381
## uniqueness.fit 55 10580 10700 84.199     17.818       8    0.02264 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ΔCFI and $\Delta\chi^2$ are significant. AIC indicates better fit for the partial scalar invariance model; BIC indicates better fit for the uniqueness invariance model.

```
lavTestScore(uniqueness.fit)$uni
```

```
##
## univariate score tests:
##
##      lhs op   rhs    X2 df p.value
## 1   .p2. == .p31. 3.973  1   0.046
## 2   .p3. == .p32. 2.156  1   0.142
## 3   .p4. == .p33. 1.200  1   0.273
## 4   .p7. == .p36. 0.084  1   0.772
## 5   .p8. == .p37. 0.872  1   0.350
## 6  .p13. == .p42. 0.028  1   0.868
## 7  .p14. == .p43. 0.200  1   0.654
## 8  .p15. == .p44. 0.241  1   0.623
## 9  .p16. == .p45. 7.287  1   0.007
## 10 .p17. == .p46. 6.339  1   0.012
## 11 .p18. == .p47. 0.055  1   0.814
## 12 .p19. == .p48. 0.010  1   0.920
## 13 .p20. == .p49. 0.368  1   0.544
## 14 .p25. == .p54. 3.123  1   0.077
## 15 .p26. == .p55. 1.491  1   0.222
## 16 .p28. == .p57. 0.186  1   0.666
## 17 .p29. == .p58. 1.085  1   0.298
```

Equality restrictions on p16 and p17 seem most problematic.

```
pars <- parameterestimates(uniqueness.fit)
pars[pars$label %in% c(".p16.", ".p17."), 1:3]
```

```
##        lhs op     rhs
## 16    Comp ~~    Comp
## 17 PicComp ~~ PicComp
## 45    Comp ~~    Comp
## 46 PicComp ~~ PicComp
```

The residual variances of Picture Completion and Comprehension do not seem equal across groups. The residual variance of the Comprehension subtest appears the worst offender of invariance, so we lift that restriction:

```
uniqueness.fit2 <- cfa(wisc3.model, sample.cov = combined.cov,
                       sample.nobs = combined.n,
                       sample.mean = combined.means,
                       group.equal = c("loadings", "intercepts", "residuals"),
                       group.partial = c("Sim~1", "VS=~PicArr", "PicArr~1",
                                         "Comp~~Comp"))
fitMeasures(uniqueness.fit2, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##    75.167    54.000     0.030     0.979     0.053     0.055 10573.116 10696.820
```

```
lavTestLRT(scalar.fit3, uniqueness.fit2)
```

```
## Chi-Squared Difference Test
```

```
##
##                  Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## scalar.fit3     47 10578 10728 66.381
## uniqueness.fit2 54 10573 10697 75.167      8.786       7     0.2684
```

We have obtained an adequately fitting partial uniqueness invariance model, according to all fit indices. We now proceed to test structural invariance across the two groups:

Structural invariance

```r
factor.var.fit <- cfa(wisc3.model, sample.cov = combined.cov,
                      sample.nobs = combined.n,
                      sample.mean = combined.means,
                      group.equal = c("loadings", "intercepts", "residuals",
                                      "lv.variances"),
                      group.partial = c("Sim~1", "VS=~PicArr", "PicArr~1",
                                        "Comp~~Comp"))
fitMeasures(factor.var.fit, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##    79.492    56.000     0.021     0.977     0.055     0.071 10573.441 10689.869
```

```r
lavTestLRT(factor.var.fit, uniqueness.fit2)
```

```
## Chi-Squared Difference Test
##
##                 Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## uniqueness.fit2 54 10573 10697 75.167
## factor.var.fit  56 10573 10690 79.492     4.3249       2      0.115
```

Equality of factor variances appears tenable. We proceed to test the equality of factor covariances:

```r
factor.covar.fit <- cfa(wisc3.model, sample.cov = combined.cov,
                        sample.nobs = combined.n,
                        sample.mean = combined.means,
                        group.equal = c("loadings", "intercepts", "residuals",
                                        "lv.variances", "lv.covariances"),
                        group.partial = c("Sim~1", "VS=~PicArr", "PicArr~1",
                                          "Comp~~Comp"))
fitMeasures(factor.covar.fit, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##    80.205    57.000     0.023     0.977     0.054     0.071 10572.154 10684.943
```

```r
lavTestLRT(factor.var.fit, factor.covar.fit)
```

```
## Chi-Squared Difference Test
##
##                  Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## factor.var.fit   56 10573 10690 79.492
## factor.covar.fit 57 10572 10685 80.205     0.7128       1     0.3985
```

Equality of the factor covariance(s) appears tenable also. Thus, manic patients do not seem to show more or less variation in intelligence than the norming group, and the association between the two latent factors seems equally strong in the two groups. We proceed to test the equality of factor means:

```r
factor.means.fit <- cfa(wisc3.model, sample.cov = combined.cov,
                        sample.nobs = combined.n,
                        sample.mean = combined.means,
                        group.equal = c("loadings", "intercepts", "residuals",
                                        "lv.variances", "lv.covariances",
                                        "means"),
                        group.partial = c("Sim~1", "VS=~PicArr", "Comp~~Comp",
                                          "PicArr~1"))
fitMeasures(factor.means.fit, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##    82.173    59.000     0.025     0.977     0.053     0.074 10570.122 10675.635
```

```r
lavTestLRT(factor.covar.fit, factor.means.fit)
```

```
## Chi-Squared Difference Test
##
##                  Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## factor.covar.fit 57 10572 10685 80.205
## factor.means.fit 59 10570 10676 82.173     1.9682       2     0.3738
```

Equality of factor means is also tenable. Thus, manic patients do not seem to have higher or lower intelligence than the norming group.

Conclusion

We obtained a well-fitting model of partial measurement invariance, and full structural invariance. This indicates that the WISc subscales measure the construct (Intelligence) in roughly a similar manner among manic persons and the norming group (probably general population). Thus, observed subscale scores can be compared between these groups. However, we also found three exceptions to measurement invariance:

```r
pars <- parameterestimates(factor.means.fit, standardized = TRUE)
pars[pars$label == "" & pars$se > 0, c(1:3, 5:8, 14)]
```

```
##       lhs op    rhs group label    est    se std.all
## 6      VS =~ PicArr     1        1.265 0.201   0.671
## 16   Comp ~~   Comp     1        2.505 0.473   0.369
## 24    Sim ~1            1        0.786 0.768   0.255
## 27 PicArr ~1            1       -2.100 2.110  -0.534
## 35     VS =~ PicArr     2        0.740 0.123   0.467
## 45   Comp ~~   Comp     2        4.824 0.534   0.530
## 53    Sim ~1            2       -0.772 0.751  -0.250
## 56 PicArr ~1            2        2.521 1.295   0.763
```

- The loading for Picture Arrangement is higher in the manic group. This subtest may be a somewhat stronger indicator of the Visuo-Spatial factor in the manic group, than in the norm group.
- The intercept for Picture Arrangement is higher for the norm group. Given the same value of the Verbal Comprehension factor, a manic person is expected to score lower than a person from the norm group on this subscale.
- The intercept for Similarities is higher in the manic group. Given the same value of the Verbal Comprehension factor, a manic person is expected to score higher than a person from the norm group on this subscale.
- The intercept differences have opposite directions, so will likely cancel out when intelligence scores are computed. But the differences should be taken into account when comparing subtest scores between

persons with and without manic symptoms.

- The residual variance for Comprehension is higher in the norm than in the manic group, indicating lower reliability for this subtest in the norm group.

Furthermore, we found that the means and (co)variances of the Verbal Comprehension and Visuo-Spatial factors did not differ between the manic and norm group. This indicates that the groups do not differ in terms of their average levels of intelligence, nor in terms of inter-individual variation, nor in the association between the two intelligence factors.

Sensitivity analysis

We perform a sensitivity analysis to see whether we would have reached a different conclusion when adding a correlated error between Block Design and Object Assembly. Adding makes sense from a substantial point of view: Both subtests require manipulation of physical objects, while the two other indicators of the Visuo-Spatial factor both involve Pictures.

```
wisc3.model2 <- '
  VC =~ Info + Sim + Vocab + Comp
  VS =~ PicComp + PicArr + BlkDsgn + ObjAsmb
  BlkDsgn ~~ ObjAsmb
  VC ~ NA*1
  VS ~ NA*1
  Info ~ 0*1
  PicComp ~ 0*1
'
configural.fit <- cfa(wisc3.model2, sample.cov = combined.cov,
                      sample.nobs = combined.n, sample.mean = combined.means,
                      meanstructure = TRUE)
fitMeasures(configural.fit, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##    39.606    36.000     0.312     0.996     0.027     0.026 10573.555 10762.749
```

The configural invariance model fits somewhat better with the correlated error. Let's inspect the value of the residual variance:

```
pars <- parameterestimates(configural.fit, standardized = TRUE)
pars[pars$op == "~~" & pars$lhs == "BlkDsgn" & pars$rhs == "ObjAsmb", c(1:3, 5:7, 9, 13)]
```

```
##        lhs op     rhs group   est    se pvalue std.all
## 9  BlkDsgn ~~ ObjAsmb     1 2.247 0.914  0.014   0.364
## 39 BlkDsgn ~~ ObjAsmb     2 1.567 0.639  0.014   0.255
```

The residual variance is significant in both groups. The estimated values differ somewhat, but the standard errors do not indicate a significant difference.

```
metric.fit <- cfa(wisc3.model2, sample.cov = combined.cov,
                  sample.nobs = combined.n, sample.mean = combined.means,
                  meanstructure = TRUE, group.equal = "loadings")
fitMeasures(metric.fit, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##    51.401    42.000     0.152     0.991     0.040     0.050 10573.350 10740.714
```

```
lavTestLRT(metric.fit, configural.fit)
```

```
## Chi-Squared Difference Test
```

```
##
##                 Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## configural.fit 36 10574 10763 39.606
## metric.fit     42 10573 10741 51.401     11.795       6     0.0667 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lavTestScore(metric.fit)
```

```
## $test
##
## total score test:
##
##    test     X2 df p.value
## 1 score 11.665  6    0.07
##
## $uni
##
## univariate score tests:
##
##    lhs op   rhs    X2 df p.value
## 1 .p2. == .p32. 2.580  1   0.108
## 2 .p3. == .p33. 0.640  1   0.424
## 3 .p4. == .p34. 0.948  1   0.330
## 4 .p6. == .p36. 5.729  1   0.017
## 5 .p7. == .p37. 0.176  1   0.675
## 6 .p8. == .p38. 0.004  1   0.952
```

```
pars <- parameterestimates(metric.fit)
pars[pars$label == ".p6.", 1:3]
```

```
##    lhs op    rhs
## 6   VS =~ PicArr
## 36  VS =~ PicArr
```

Metric invariance is tenable according to BIC, AIC, $\Delta\chi^2$ and ΔCFI. However, the most invariance-violating parameter is still the loading of Picture Arrangement on the VS factor; this is similar to what we found in our earlier analysis. I will ift this restriction and continue testing scalar invariance:

```
scalar.fit <- cfa(wisc3.model2, sample.cov = combined.cov,
                  sample.nobs = combined.n, sample.mean = combined.means,
                  meanstructure = TRUE,
                  group.equal = c("loadings", "intercepts"),
                  group.partial = c("VS=~PicArr"))
fitMeasures(scalar.fit, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##    89.217    47.000     0.000     0.959     0.080     0.056 10601.166 10750.339
```

```
lavTestLRT(metric.fit, scalar.fit)
```

```
## Chi-Squared Difference Test
##
##             Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## metric.fit 42 10573 10741 51.401
## scalar.fit 47 10601 10750 89.217     37.817       5  4.107e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
lavTestScore(scalar.fit)$uni
```

```
##
## univariate score tests:
##
##      lhs op   rhs     X2 df p.value
## 1   .p2. == .p32. 31.021  1   0.000
## 2   .p3. == .p33.  0.966  1   0.326
## 3   .p4. == .p34.  6.134  1   0.013
## 4   .p7. == .p37.  0.072  1   0.789
## 5   .p8. == .p38.  0.902  1   0.342
## 6  .p25. == .p55. 29.567  1   0.000
## 7  .p26. == .p56.  0.792  1   0.373
## 8  .p27. == .p57.  7.675  1   0.006
## 9  .p28. == .p58.  4.576  1   0.032
## 10 .p29. == .p59.  0.043  1   0.836
## 11 .p30. == .p60.  0.930  1   0.335
```

```r
pars <- parameterestimates(scalar.fit)
pars[pars$label %in% c(".p2.", ".p25."), 1:3]
```

```
##    lhs op rhs
## 2   VC =~ Sim
## 25 Sim ~1
## 32  VC =~ Sim
## 55 Sim ~1
```

Scalar invariance is not tenable, like in our earlier analysis. Also, the intercept of Similarities seems to be the most violating parameter, like in our earlier analysis. We lift the equality restriction on the intercept of the Similarities subtest:

```r
scalar.fit2 <- cfa(wisc3.model2, sample.cov = combined.cov,
                   sample.nobs = combined.n, sample.mean = combined.means,
                   group.equal = c("loadings", "intercepts"),
                   group.partial = c("VS=~PicArr", "Sim~1"),
                   meanstructure = TRUE)
fitMeasures(scalar.fit2, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##    55.970    46.000     0.149     0.990     0.039     0.049 10569.919 10722.730
```

```r
lavTestLRT(metric.fit, scalar.fit2)
```

```
## Chi-Squared Difference Test
##
##             Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## metric.fit  42 10573 10741 51.401
## scalar.fit2 46 10570 10723 55.970      4.569       4     0.3344
```

We obtained partial scalar invariance, like in the earlier analysis. We continuous testing uniqueness invariance:

```r
uniqueness.fit <- cfa(wisc3.model2, sample.cov = combined.cov,
                      sample.nobs = combined.n,
                      sample.mean = combined.means,
                      group.equal=c("loadings", "intercepts", "residuals"),
                      group.partial = c("VS=~PicArr", "Sim~1"))
fitMeasures(uniqueness.fit, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##    76.514    54.000     0.024     0.978     0.054     0.060 10574.463 10698.167
```

```r
lavTestLRT(uniqueness.fit, scalar.fit2)
```

```
## Chi-Squared Difference Test
##
##                Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## scalar.fit2    46 10570 10723 55.970
## uniqueness.fit 54 10574 10698 76.514     20.544       8   0.008462 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
lavTestScore(uniqueness.fit)$uni
```

```
##
## univariate score tests:
##
##      lhs op   rhs    X2 df p.value
## 1   .p2. == .p32. 4.028  1   0.045
## 2   .p3. == .p33. 2.164  1   0.141
## 3   .p4. == .p34. 1.224  1   0.269
## 4   .p7. == .p37. 0.011  1   0.915
## 5   .p8. == .p38. 0.646  1   0.422
## 6  .p14. == .p44. 0.019  1   0.891
## 7  .p15. == .p45. 0.169  1   0.681
## 8  .p16. == .p46. 0.224  1   0.636
## 9  .p17. == .p47. 7.441  1   0.006
## 10 .p18. == .p48. 7.645  1   0.006
## 11 .p19. == .p49. 1.280  1   0.258
## 12 .p20. == .p50. 0.048  1   0.826
## 13 .p21. == .p51. 0.324  1   0.569
## 14 .p26. == .p56. 3.140  1   0.076
## 15 .p27. == .p57. 1.517  1   0.218
## 16 .p28. == .p58. 6.153  1   0.013
## 17 .p29. == .p59. 0.021  1   0.886
## 18 .p30. == .p60. 0.688  1   0.407
```

```r
pars <- parameterestimates(uniqueness.fit)
pars[pars$label %in% c(".p17.", ".p18."), 1:3]
```

```
##        lhs op     rhs
## 17    Comp ~~    Comp
## 18 PicComp ~~ PicComp
## 47    Comp ~~    Comp
## 48 PicComp ~~ PicComp
```

Again, residual variance of Comprehension and Picture Completion appear to be the most invariance-violating parameters. Picture Completion seems a slightly stronger violator than Comprehesion; in our earlier analysis this was the other way around, but the difference is slight. Let's lift the restriction on both, so we can continue evaluating structural invariance:

```r
uniqueness.fit2 <- cfa(wisc3.model2, sample.cov = combined.cov,
                       sample.nobs = combined.n,
                       sample.mean = combined.means,
                       group.equal = c("loadings", "intercepts", "residuals"),
                       group.partial = c("Sim~1", "VS=~PicArr",
```

```r
                                          "PicComp~~PicComp", "Comp ~~ Comp"))
fitMeasures(uniqueness.fit2, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##    57.974    52.000     0.264     0.994     0.029     0.053 10559.923 10690.904
```

```r
lavTestLRT(scalar.fit2, uniqueness.fit2)
```

```
## Chi-Squared Difference Test
##
##                 Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## scalar.fit2     46 10570 10723 55.970
## uniqueness.fit2 52 10560 10691 57.974     2.0044       6     0.9193
```

I will restict the structural parameters to equality in one single step:

```r
structural.fit <- cfa(wisc3.model2, sample.cov = combined.cov,
                      sample.nobs = combined.n,
                      sample.mean = combined.means,
                      group.equal = c("loadings", "intercepts", "residuals",
                                      "lv.variances", "lv.covariances",
                                      "means"),
                      group.partial = c("VS=~PicArr", "Sim~1",
                                        "PicComp~~PicComp", "Comp ~~ Comp"))
fitMeasures(structural.fit, fit.indices)
```

```
##     chisq        df    pvalue       cfi     rmsea      srmr       aic       bic
##    65.226    57.000     0.212     0.992     0.032     0.068 10557.175 10669.964
```

```r
lavTestLRT(structural.fit, uniqueness.fit2)
```

```
## Chi-Squared Difference Test
##
##                 Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## uniqueness.fit2 52 10560 10691 57.974
## structural.fit  57 10557 10670 65.226     7.2518       5     0.2026
```

Structural invariance is tenable, too. Note that the Δdf value of 5 comes from applying equality restrictions on two LV means, two LV variances and 1 LV covariance. This yields 5 additional degrees of freedom.

Conclusion Sensitivity Analyses

After adding a correlated error to the model, we reach very similar conclusions: Most measurement parameters are equal between groups, all structural parameters are equal between groups. Variance-offending parameters appeared the same between the model with and without correlated errors.