# A Tutorial on Hierarchically Structured Constructs

## Online Supplement

**Martin Brunner**
University of Luxembourg
Luxembourg

**Gabriel Nagy**
University of Tuebingen
Germany

**Oliver Wilhelm**
University of Ulm
Germany

Contact: martin.brunner@uni.lu

Online supplement to the article:

Brunner, M., Nagy, G., & Wilhelm, O. (in press). A tutorial on hierarchically structured constructs. *Journal of Personality*.

The first part of this online supplement contains the Mplus input files (Muthén & Muthén, 1998–2010) that we used to analyze the data of the Spanish standardization sample of the WAIS–III (Colom, Abad, Garcia, & Juan-Espinosa, 2002, Table A3) by means of the one-factor model, the first-order factor model, the higher-order factor model, and the nested-factor model. The second part contains the Mplus input files and R code (R Development Core Team, 2011) that we used to implement the multi-step procedure developed by Stoel and colleagues (Stoel, Garre, Dolan, & van den Wittenboer, 2006) to test the first-order factor model against the one-factor model.

## Mplus Input Files Used to Analyze the CFA models

To analyze the various CFA models, we entered Table 1 into a text file (i.e., wais3_spain.txt) in the order presented. Note that the model specification of all four CFA models takes into account that we analyzed a correlation matrix rather than a covariance matrix (see Footnote 2 in the article). To this end, we introduced latent phantom variables that rescaled the variance of each manifest subtest score to 1.

### *One-Factor Model*

```
TITLE: ONE-FACTOR MODEL

DATA:
!specification of the input file that needs to be located in the same folder
!as this input file
FILE IS wais3_spain.txt;

!indicates that the input file is a correlation matrix
TYPE=CORR;

!indicates that the correlations are based on data from 1,369 persons
NOBSERVATION ARE 1369;

VARIABLE:
!names of the variables of the corresponding correlation table
!the variables must appear in the same order as they appear in the correlation matrix
NAMES ARE voc sim ari dig inf com let pic_c cod blo mat pic_a sym obj;

!names of the variables that are used in the CFA models
USEVARIABLES ARE voc sim ari dig inf com let pic_c cod blo mat pic_a sym obj;

ANALYSIS:
!we require the standard structural equation framework as the latent variable environment
TYPE=GENERAL;

MODEL:
!To facilitate convergence of the model estimation procedure, we provided
!starting values for the factor loadings on g

!Phantom variables (starting with i) are needed to ensure that the variance
!of the manifest variables is exactly 1

ivoc by voc*;
isim by sim*;
iari by ari*;
idig by dig*;
iinf by inf*;
icom by com*;
ilet by let*;
ipic_c by pic_c*;
icod by cod*;
```

```
iblo by blo*;
imat by mat*;
ipic_a by pic_a*;
isym by sym*;
iobj by obj*;

!the residual variance of the manifest variables is fixed to zero.
!Thus, the total variance of the manifest variables is contained in the phantom variables

voc@0;
sim@0;
ari@0;
dig@0;
inf@0;
com@0;
let@0;
pic_c@0;
cod@0;
blo@0;
mat@0;
pic_a@0;
sym@0;
obj@0;

!the variances of the residual terms of the phantom variables will be
!constrained to equal “1 - squared loading on g” (see “MODEL CONSTRAINT” section).

ivoc   (e1)
isim   (e2)
iari   (e3)
idig   (e4)
iinf   (e5)
icom   (e6)
ilet   (e7)
ipic_c (e8)
icod   (e9)
iblo   (e10)
imat   (e11)
ipic_a (e12)
isym   (e13)
iobj   (e14);

!General Cognitive Ability
!starting values (i.e., *.8) are provided to facilitate
!the convergence of the parameter estimation
!the factor loadings are given labels (la1 to la14)
!these labels are needed to compute score reliability as proposed by Cheung (2009)
g by ivoc*.8   (la1)
     isim*.8   (la2)
     iari*.8   (la3)
     idig*.8   (la4)
     iinf*.8   (la5)
     icom*.8   (la6)
     ilet*.8   (la7)
     ipic_c*.8 (la8)
     icod*.8   (la9)
     iblo*.8   (la10)
     imat*.8   (la11)
     ipic_a*.8 (la12)
     isym*.8   (la13)
     iobj*.8   (la14);

!the one-factor model is identified by fixing the latent variance of g_OF to 1
g@1;

MODEL CONSTRAINT:
!constraining the error variances of the phantom variables to “1 – variance
!explained by g”;
!the error terms are entered in the computation of score reliability

e1 = 1 - la1^2;
e2 = 1 - la2^2;
e3 = 1 - la3^2;
e4 = 1 - la4^2;
e5 = 1 - la5^2;
e6 = 1 - la6^2;
e7 = 1 - la7^2;
e8 = 1 - la8^2;
```

```
e9 = 1 - la9^2;
e10 = 1 - la10^2;
e11 = 1 - la11^2;
e12 = 1 - la12^2;
e13 = 1 - la13^2;
e14 = 1 - la14^2;


!computation of omega (= o) of the General Cognitive Ability score (see Cheung, 2009)
!To this end, a new variable (o_g) is introduced

new(o_g);

!o_g is then computed as described in the manuscript

o_g = (la1+la2+la3+la4+la5+la6+la7+la8+la9+la10+la11+la12+la13+la14)^2/
      ((la1+la2+la3+la4+la5+la6+la7+la8+la9+la10+la11+la12+la13+la14)^2 +
       e1+e2+e3+e4+e5+e6+e7+e8+e9+e10+e11+e12+e13+e14);

OUTPUT: STANDARDIZED; !this requests standardized model parameters
```

*First-Order Factor Model*

```
TITLE: First-Order Factor Model

DATA:
FILE IS wais3_spain.txt;
TYPE=CORR;
NOBSERVATION ARE 1369;

VARIABLE:
NAMES ARE voc sim ari dig inf com let pic_c cod blo mat pic_a sym obj;
USEVARIABLES ARE voc sim ari dig inf com let pic_c cod blo mat pic_a sym obj;

ANALYSIS:
TYPE=GENERAL ;

MODEL:
!To facilitate convergence of the model estimation procedure, we provided starting values

!Phantom Variables
ivoc by voc*1;
isim by sim*1;
iari by ari*1;
idig by dig*1;
iinf by inf*1;
icom by com*1;
ilet by let*1;
ipic_c by pic_c*1;
icod by cod*1;
iblo by blo*1;
imat by mat*1;
ipic_a by pic_a*1;
isym by sym*1;
iobj by obj*1;

voc@0;
sim@0;
ari@0;
dig@0;
inf@0;
com@0;
let@0;
pic_c@0;
cod@0;
blo@0;
mat@0;
pic_a@0;
sym@0;
obj@0;

iinf*.36   (e1);
ivoc*.36   (e2);
isim*.36   (e3);
icom*.36   (e4);
iobj*.36   (e5);
iblo*.36   (e6);
ipic_c*.36 (e7);
imat*.36   (e8);
ipic_a*.36 (e9);
idig*.36   (e10);
ilet*.36   (e11);
iari*.36   (e12);
icod*.36   (e13);
isym*.36   (e14);

!Verbal Comprehension
vc by iinf*.8  (la1)
      ivoc*.8  (la2)
      isim*.8  (la3)
      icom*.8  (la4);

vc@1;

!Perceptual Organization
po by iobj*.8   (la5)
      iblo*.8   (la6)
      ipic_c*.8 (la7)
      imat*.8   (la8)
```

```
      ipic_a*.8 (la9);

po@1;

!Working Memory
wm by idig*.8   (la10)
      ilet*.8   (la11)
      iari*.8   (la12) ;

wm@1;

!Processing Speed
ps by icod*.8  (la13)
      isym*.8  (la14);

ps@1;

MODEL CONSTRAINT:

e1 = 1 - la1^2;
e2 = 1 - la2^2;
e3 = 1 - la3^2;
e4 = 1 - la4^2;
e5 = 1 - la5^2;
e6 = 1 - la6^2;
e7 = 1 - la7^2;
e8 = 1 - la8^2;
e9 = 1 - la9^2;
e10 = 1 - la10^2;
e11 = 1 - la11^2;
e12 = 1 - la12^2;
e13 = 1 - la13^2;
e14 = 1 - la14^2;

new(o_vc o_po o_wm o_ps);

o_vc = (la1+la2+la3+la4)^2/((la1+la2+la3+la4)^2 + e1+e2+e3+e4);

o_po = (la5+la6+la7+la8+la9)^2/((la5+la6+la7+la8+la9)^2 + e5+e6+e7+e8+e9);

o_wm = (la10+la11+la12)^2/((la10+la11+la12)^2 + e10+e11+e12);

o_ps = (la13+la14)^2/((la13+la14)^2 + e13+e14);

OUTPUT: STANDARDIZED;
```

# *Higher-Order Factor Model*

```
TITLE: Higher-Order Factor Model

DATA:
FILE IS wais3_spain.txt;
TYPE=CORR;
NOBSERVATION ARE 1369;

VARIABLE:
NAMES ARE voc sim ari dig inf com let pic_c cod blo mat pic_a sym obj;
USEVARIABLES ARE voc sim ari dig inf com let pic_c cod blo mat pic_a sym obj;

ANALYSIS:
TYPE=GENERAL ;

MODEl:

!To facilitate convergence of the model estimation procedure, we provided starting values

!Phantom Variables
ivoc by voc*1;
isim by sim*1;
iari by ari*1;
idig by dig*1;
iinf by inf*1;
icom by com*1;
ilet by let*1;
ipic_c by pic_c*1;
icod by cod*1;
iblo by blo*1;
imat by mat*1;
ipic_a by pic_a*1;
isym by sym*1;
iobj by obj*1;

voc@0;
sim@0;
ari@0;
dig@0;
inf@0;
com@0;
let@0;
pic_c@0;
cod@0;
blo@0;
mat@0;
pic_a@0;
sym@0;
obj@0;

iinf*.36   (e1);
ivoc*.36   (e2);
isim*.36   (e3);
icom*.36   (e4);
iobj*.36   (e5);
iblo*.36   (e6);
ipic_c*.36 (e7);
imat*.36   (e8);
ipic_a*.36 (e9);
idig*.36   (e10);
ilet*.36   (e11);
iari*.36   (e12);
icod*.36   (e13);
isym*.36   (e14);

!Verbal Comprehension
vc by iinf*.8  (la1)
      ivoc*.8  (la2)
      isim*.8  (la3)
      icom*.8  (la4);

vc (var_vc);

!Perceptual Organization
po by iobj*.8   (la5)
      iblo*.8   (la6)
      ipic_c*.8 (la7)
```

```
      imat*.8   (la8)
      ipic_a*.8 (la9);

po (var_po);

!Working Memory
wm by idig*.8 (la10)
      ilet*.8 (la11)
      iari*.8 (la12) ;

wm (var_wm);

!Processing Speed
ps by icod*.8 (la13)
      isym*.8 (la14);

ps (var_ps);

!Genenearal Cognitive Ability
g by  vc*.8  (la15)
      po*.8  (la16)
      wm*.8  (la17)
      ps*.8  (la18);

g@1;

MODEL CONSTRAINT:

e1 = 1 - la1^2;
e2 = 1 - la2^2;
e3 = 1 - la3^2;
e4 = 1 - la4^2;
e5 = 1 - la5^2;
e6 = 1 - la6^2;
e7 = 1 - la7^2;
e8 = 1 - la8^2;
e9 = 1 - la9^2;
e10 = 1 - la10^2;
e11 = 1 - la11^2;
e12 = 1 - la12^2;
e13 = 1 - la13^2;
e14 = 1 - la14^2;

!the variance of the residualized first-order factors is
!equal to 1 - the variance that is explained by gHO
!(i.e., the square of the corresponding factor loading on gHO)

var_vc = 1 - la15^2;
var_po = 1 - la16^2;
var_wm = 1 - la17^2;
var_ps = 1 - la18^2;

!omega = o; omega hierarcical = oH
new(o_vc o_po o_wm o_ps o_g oH_vc oH_po oH_wm oH_ps oH_g);

!The multiplications represent the transformation of factor loadings
!according to the Schmid-Leiman transformation
o_vc = ((la1*la15+la2*la15+la3*la15+la4*la15)^2 +
        (la1*sqrt(var_vc)+la2*sqrt(var_vc)+la3*sqrt(var_vc)+la4*sqrt(var_vc))^2) /
       ((la1*la15+la2*la15+la3*la15+la4*la15)^2 +
        (la1*sqrt(var_vc)+la2*sqrt(var_vc)+la3*sqrt(var_vc)+la4*sqrt(var_vc))^2 +
        e1+e2+e3+e4);

o_po = ((la5*la16+la6*la16+la7*la16+la8*la16+la9*la16)^2 +
         (la5*sqrt(var_po)+la6*sqrt(var_po)+la7*sqrt(var_po)+
         la8*sqrt(var_po)+la9*sqrt(var_po))^2) /
       ((la5*la16+la6*la16+la7*la16+la8*la16+la9*la16)^2 +
         (la5*sqrt(var_po)+la6*sqrt(var_po)+la7*sqrt(var_po)+
         la8*sqrt(var_po)+la9*sqrt(var_po))^2 +
         e5+e6+e7+e8+e9);

o_wm = ((la10*la17+la11*la17+la12*la17)^2 +
        (la10*sqrt(var_wm)+la11*sqrt(var_wm)+la12*sqrt(var_wm))^2) /
       ((la10*la17+la11*la17+la12*la17)^2 +
        (la10*sqrt(var_wm)+la11*sqrt(var_wm)+la12*sqrt(var_wm))^2 +
        e10+e11+e12);

o_ps = ((la13*la18+la14*la18)^2 +
```

```
        (la13*sqrt(var_ps)+la14*sqrt(var_ps))^2) /
       ((la13*la18+la14*la18)^2 +
        (la13*sqrt(var_ps)+la14*sqrt(var_ps))^2 +
        e13+e14);

o_g  = ((la1*la15+la2*la15+la3*la15+la4*la15 +
          la5*la16+la6*la16+la7*la16+la8*la16+la9*la16 +
          la10*la17+la11*la17+la12*la17 +
          la13*la18+la14*la18)^2 +
         (la1*sqrt(var_vc)+la2*sqrt(var_vc)+la3*sqrt(var_vc)+la4*sqrt(var_vc))^2 +
         (la5*sqrt(var_po)+la6*sqrt(var_po)+la7*sqrt(var_po)+
          la8*sqrt(var_po)+la9*sqrt(var_po))^2 +
         (la10*sqrt(var_wm)+la11*sqrt(var_wm)+la12*sqrt(var_wm))^2 +
         (la13*sqrt(var_ps)+la14*sqrt(var_ps))^2) /
        ((la1*la15+la2*la15+la3*la15+la4*la15 +
          la5*la16+la6*la16+la7*la16+la8*la16+la9*la16 +
          la10*la17+la11*la17+la12*la17 +
          la13*la18+la14*la18)^2 +
         (la1*sqrt(var_vc)+la2*sqrt(var_vc)+la3*sqrt(var_vc)+la4*sqrt(var_vc))^2 +
         (la5*sqrt(var_po)+la6*sqrt(var_po)+la7*sqrt(var_po)+
          la8*sqrt(var_po)+la9*sqrt(var_po))^2 +
         (la10*sqrt(var_wm)+la11*sqrt(var_wm)+la12*sqrt(var_wm))^2 +
         (la13*sqrt(var_ps)+la14*sqrt(var_ps))^2 +
          e1+e2+e3+e4+e5+e6+e7+e8+e9+e10+e11+e12+e13+e14);

oH_vc = ((la1*sqrt(var_vc)+la2*sqrt(var_vc)+la3*sqrt(var_vc)+la4*sqrt(var_vc))^2) /
        ((la1*la15+la2*la15+la3*la15+la4*la15)^2 +
         (la1*sqrt(var_vc)+la2*sqrt(var_vc)+la3*sqrt(var_vc)+la4*sqrt(var_vc))^2 +
         e1+e2+e3+e4);

oH_po = ((la5*sqrt(var_po)+la6*sqrt(var_po)+la7*sqrt(var_po)+
         la8*sqrt(var_po)+la9*sqrt(var_po))^2) /
        ((la5*la16+la6*la16+la7*la16+la8*la16+la9*la16)^2 +
         (la5*sqrt(var_po)+la6*sqrt(var_po)+la7*sqrt(var_po)+
          la8*sqrt(var_po)+la9*sqrt(var_po))^2 +
         e5+e6+e7+e8+e9);

oH_wm = ((la10*sqrt(var_wm)+la11*sqrt(var_wm)+la12*sqrt(var_wm))^2) /
       ((la10*la17+la11*la17+la12*la17)^2 +
         (la10*sqrt(var_wm)+la11*sqrt(var_wm)+la12*sqrt(var_wm))^2 +
         e10+e11+e12);

oH_ps = ((la13*sqrt(var_ps)+la14*sqrt(var_ps))^2) /
        ((la13*la18+la14*la18)^2 +
         (la13*sqrt(var_ps)+la14*sqrt(var_ps))^2 +
         e13+e14);

oH_g  = ((la1*la15+la2*la15+la3*la15+la4*la15 +
          la5*la16+la6*la16+la7*la16+la8*la16+la9*la16 +
          la10*la17+la11*la17+la12*la17 +
          la13*la18+la14*la18)^2)  /
         ((la1*la15+la2*la15+la3*la15+la4*la15 +
           la5*la16+la6*la16+la7*la16+la8*la16+la9*la16 +
           la10*la17+la11*la17+la12*la17 +
           la13*la18+la14*la18)^2 +
         (la1*sqrt(var_vc)+la2*sqrt(var_vc)+la3*sqrt(var_vc)+la4*sqrt(var_vc))^2 +
         (la5*sqrt(var_po)+la6*sqrt(var_po)+la7*sqrt(var_po)+
          la8*sqrt(var_po)+la9*sqrt(var_po))^2 +
         (la10*sqrt(var_wm)+la11*sqrt(var_wm)+la12*sqrt(var_wm))^2 +
         (la13*sqrt(var_ps)+la14*sqrt(var_ps))^2 +
          e1+e2+e3+e4+e5+e6+e7+e8+e9+e10+e11+e12+e13+e14);

!The Tech4 output requests the correlations among the first-order factors
!that are implied by the higher-order factor
!these model-implied correlations are subtracted from the corresponding
!correlations obtained for the first-order factor model to compute the
!residual correlations among the first-order factors

OUTPUT: STANDARDIZED TECH4;
```

# *Nested-Factor Model*

```
TITLE: Nested-Factor Model

DATA:
FILE IS wais3_spain.txt;
TYPE=CORR;
NOBSERVATION ARE 1369;

VARIABLE:
NAMES ARE voc sim ari dig inf com let pic_c cod blo mat pic_a sym obj;
USEVARIABLES ARE voc sim ari dig inf com let pic_c cod blo mat pic_a sym obj;

ANALYSIS:
TYPE=GENERAL ;

MODEl:

!To facilitate convergence of the model estimation procedure, we provided starting values

!Phantom Variables
ivoc by voc*1;
isim by sim*1;
iari by ari*1;
idig by dig*1;
iinf by inf*1;
icom by com*1;
ilet by let*1;
ipic_c by pic_c*1;
icod by cod*1;
iblo by blo*1;
imat by mat*1;
ipic_a by pic_a*1;
isym by sym*1;
iobj by obj*1;

voc@0;
sim@0;
ari@0;
dig@0;
inf@0;
com@0;
let@0;
pic_c@0;
cod@0;
blo@0;
mat@0;
pic_a@0;
sym@0;
obj@0;

iinf*.36   (e1);
ivoc*.36   (e2);
isim*.36   (e3);
icom*.36   (e4);
iobj*.36   (e5);
iblo*.36   (e6);
ipic_c*.36 (e7);
imat*.36   (e8);
ipic_a*.36 (e9);
idig*.36   (e10);
ilet*.36   (e11);
iari*.36   (e12);
icod*.36   (e13);
isym*.36   (e14);


!specific Verbal Comprehension
vc by iinf*.3   (la1)
      ivoc*.3   (la2)
      isim*.3   (la3)
      icom*.3   (la4);

vc@1;

!specific Perceptual Organization
po by iobj*.3  (la5)
      iblo*.3   (la6)
```

```
      ipic_c*.3 (la7)
      imat*.3   (la8)
      ipic_a*.3 (la9);

po@1;

!specific Working Memory
wm by idig*.3  (la10)
      ilet*.3   (la11)
      iari*.3   (la12) ;

wm@1;

!specific Processing Speed

!Preliminary analyses indicated empirical underidentification when both
!factor loadings on specific processing speed were freely identified.
!To overcome this problem, we constrained the factor loadings of cod and sym
!to be equal (i.e., both factor loadings take on the parameter value of la13)

ps by icod*.3  (la13)
      isym*.3  (la13);

ps@1;

!Geneneral Cognitive Ability
g   by  iinf*.8   (la14)
        ivoc*.8   (la15)
        isim*.8   (la16)
        icom*.8   (la17)
        iobj*.8   (la18)
        iblo*.8   (la19)
        ipic_c*.8 (la20)
        imat*.8   (la21)
        ipic_a*.8 (la22)
        idig*.8   (la23)
        ilet*.8   (la24)
        iari*.8   (la25)
        icod*.8   (la26)
        isym*.8   (la27);

g@1;

g   with vc@0 po@0 wm@0 ps@0;
vc  with po@0 wm@0 ps@0;
po  with wm@0 ps@0;
wm  with ps@0;

MODEL CONSTRAINT:
!constraining the error terms of the phantom variables to 1 - variance explained by g;

e1 = 1 - la1^2 - la14^2;
e2 = 1 - la2^2 - la15^2;
e3 = 1 - la3^2 - la16^2;
e4 = 1 - la4^2 - la17^2;
e5 = 1 - la5^2 - la18^2;
e6 = 1 - la6^2 - la19^2;
e7 = 1 - la7^2 - la20^2;
e8 = 1 - la8^2 - la21^2;
e9 = 1 - la9^2 - la22^2;
e10 = 1 - la10^2 - la23^2;
e11 = 1 - la11^2 - la24^2;
e12 = 1 - la12^2 - la25^2;
e13 = 1 - la13^2 - la26^2;
e14 = 1 - la13^2 - la27^2;

!o  = omega, oH = omega hierarchical

new(o_vc o_po o_wm o_ps o_gBF oH_vc oH_po oH_wm oH_ps oH_gBF);

o_vc = ((la1+la2+la3+la4)^2 + (la14+la15+la16+la17)^2)/
       ((la1+la2+la3+la4)^2 + (la14+la15+la16+la17)^2 + e1+e2+e3+e4);

o_po = ((la5+la6+la7+la8+la9)^2 + (la18+la19+la20+la21+la22)^2)/
       ((la5+la6+la7+la8+la9)^2 + (la18+la19+la20+la21+la22)^2 + e5+e6+e7+e8+e9);

o_wm = ((la10+la11+la12)^2 + (la23+la24+la25)^2)/
       ((la10+la11+la12)^2 + (la23+la24+la25)^2 + e10+e11+e12);
```

```
o_ps = ((la13+la13)^2 + (la26+la27)^2)/
       ((la13+la13)^2 + (la26+la27)^2 + e13+e14);

o_gBF = ((la1+la2+la3+la4)^2 +
         (la5+la6+la7+la8+la9)^2 +
         (la10+la11+la12)^2 +
         (la13+la13)^2 +
         (la14+la15+la16+la17+la18+la19+la20+la21+la22+la23+la24+la25+la26+la27)^2)/
        ((la1+la2+la3+la4)^2 +
         (la5+la6+la7+la8+la9)^2 +
         (la10+la11+la12)^2 +
         (la13+la13)^2 +
         (la14+la15+la16+la17+la18+la19+la20+la21+la22+la23+la24+la25+la26+la27)^2 +
          e1+e2+e3+e4+e5+e6+e7+e8+e9+e10+e11+e12+e13+e14);

oH_vc = (la1+la2+la3+la4)^2/
        ((la1+la2+la3+la4)^2 + (la14+la15+la16+la17)^2 + e1+e2+e3+e4);

oH_po = (la5+la6+la7+la8+la9)^2/
        ((la5+la6+la7+la8+la9)^2 + (la18+la19+la20+la21+la22)^2 + e5+e6+e7+e8+e9);

oH_wm = (la10+la11+la12)^2/
        ((la10+la11+la12)^2 + (la23+la24+la25)^2 + e10+e11+e12);

oH_ps = (la13+la13)^2/
        ((la13+la13)^2 + (la26+la27)^2 + e13+e14);

oH_gBF = (la14+la15+la16+la17+la18+la19+la20+la21+la22+la23+la24+la25+la26+la27)^2/
         ((la1+la2+la3+la4)^2 +
          (la5+la6+la7+la8+la9)^2 +
          (la10+la11+la12)^2 +
          (la13+la13)^2 +
          (la14+la15+la16+la17+la18+la19+la20+la21+la22+la23+la24+la25+la26+la27)^2 +
           e1+e2+e3+e4+e5+e6+e7+e8+e9+e10+e11+e12+e13+e14);

OUTPUT: STANDARDIZED;
```

**Mplus Input Files and R-Code to Implement the Stoel et al. (2006) procedure**

The multistep procedure by Stoel et al. (2006) has two components: (a) a simulation study by means of Mplus and (b) an analysis of some results of this simulation study by means of R.

We first consider the simulation study. To evaluate the statistical significance of the difference between $\chi^2$ goodness-of-fit statistics (when boundary parameters are involved), we used a $\bar{\chi}^2$ distribution that reflects a mixture of several $\chi^2$ distributions with varying degrees of freedom. The number of degrees of freedom of these distributions depends on the number of parameters that were fixed to their boundary values. In the present example, the numbers of degrees of freedom of the various $\chi^2$ distributions were $df = 0$, $df = 1$, $df = 2$, $df = 3$, $df = 4$, $df = 5$, and $df = 6$, corresponding to six factor correlations among four first-order factors. To compute the $\bar{\chi}^2$ distribution, we combined these $\chi^2$ distributions by weighting each individual $\chi^2$ distribution. The weights were determined by a simulation study in which the first-order factor model (Figure 1b) was fitted to 10,000 samples (with the actual sample size, i.e., $N$ = 1,369). These samples were drawn from a population in which the true model was a first-order factor model containing four first-order factors, where factor loadings and error terms were identical to those obtained for the first-order factor model (Figure 1b), but all factor correlations were fixed to $r = 1$ (this model is equivalent to a one-factor model). The next section contains the input file for the simulation study.

*Mplus Input File for the Simulation Study*

```
TITLE: Monte Carlo study to compute the weights for the
chi-bar-square distribution to test the one-factor model against
the first-order factor model for the WAIS-III data (Colom et al., 2002).
The input syntax was adapted from Stoel et al., 2006,
as retrieved on March 10, 2011 from
http://dx.doi.org/10.1037/1082-989X.11.4.439.supp

MONTECARLO:
!definition of the variable names
names are voc sim ari dig inf com let pic_c cod blo mat pic_a sym obj;

!number of observations in each sample (this is the sample size of Colom et al., 2002)
nobservations = 1369;

!number of replications (set to 10,000 as recommended by Stoel et al., 2006)
nreps = 10000;
```

```
!set the seed for random data generation
seed = 12345;

!file containing parameter estimates and likelihood values of each sample
!this file needs to be further processed to compute the weights
results = stoel_test.res;

!The following lines contain the specification of the population model
!The population values of the factor loadings and residual terms are fixed
!to the standardized values as obtained for the first-order factor model (Figure 1b).

MODEL POPULATION:


VC BY
   INF@0.821
   VOC@0.880
   SIM@0.854
   COM@0.810;

PO BY
   OBJ@0.796
   BLO@0.854
   PIC_C@0.8
   MAT@0.894
   PIC_A@0.839;

WM BY
   DIG@0.811
   LET@0.896
   ARI@0.78;

PS BY
   COD@0.889
   SYM@0.886;

!Residual Variances
VOC@0.225;
SIM@0.270;
ARI@0.391;
DIG@0.343;
INF@0.327;
COM@0.344;
LET@0.198;
PIC_C@0.360;
COD@0.210;
BLO@0.271;
MAT@0.200;
PIC_A@0.296;
SYM@0.216;
OBJ@0.366;


!Factor variances are set to 1 to identify the model. Doing so also facilitates
!the specification of all four factors to collapse into one single factor
VC@1;
PO@1;
WM@1;
PS@1;


!All factor correlations are set to 1 (i.e., this is equivalent to a one-factor model)

PO WITH VC@1.000;

WM WITH VC@1.000 PO@1.000;

PS WITH VC@1.000 PO@1.000 WM@1.000;


ANALYSIS:
type=general ;


MODEl:

!creating phantom latent variables that represent manifest variables
!with total variances rescaled to 1
```

```
!reasonable starting values are provided to allow convergence of the model
!starting values are 1.0 (this is the expected scaling factor)


ivoc by voc*1;
isim by sim*1;
iari by ari*1;
idig by dig*1;
iinf by inf*1;
icom by com*1;
ilet by let*1;
ipic_c by pic_c*1;
icod by cod*1;
iblo by blo*1;
imat by mat*1;
ipic_a by pic_a*1;
isym by sym*1;
iobj by obj*1;

voc@0;
sim@0;
ari@0;
dig@0;
inf@0;
com@0;
let@0;
pic_c@0;
cod@0;
blo@0;
mat@0;
pic_a@0;
sym@0;
obj@0;

!the variances of the residual terms of the phantom variables will be constrained to equal
!1 - squared loading on the respective first-order factor (see model constraint section).

iinf*.36   (e1);
ivoc*.36   (e2);
isim*.36   (e3);
icom*.36   (e4);
iobj*.36   (e5);
iblo*.36   (e6);
ipic_c*.36 (e7);
imat*.36   (e8);
ipic_a*.36 (e9);
idig*.36   (e10);
ilet*.36   (e11);
iari*.36   (e12);
icod*.36   (e13);
isym*.36   (e14);


!Verbal Comprehension
vc by iinf*.8  (la1)
      ivoc*.8  (la2)
      isim*.8  (la3)
      icom*.8  (la4);

vc@1;


!Perceptual Organization
po by iobj*.8   (la5)
      iblo*.8   (la6)
      ipic_c*.8 (la7)
      imat*.8   (la8)
      ipic_a*.8 (la9);

po@1;


!Working Memory
wm by idig*.8   (la10)
      ilet*.8   (la11)
      iari*.8   (la12) ;
```

```
wm@1;


!Processing Speed
ps by icod*.8  (la13)
      isym*.8  (la14);

ps@1;

!All factor correlations are freely estimated
!(i.e., as proposed by the alternative hypothesis to be tested)
PO WITH VC*;

WM WITH VC* PO*;

PS WITH VC* PO* WM*;


MODEL CONSTRAINT:

e1 = 1 - la1^2;
e2 = 1 - la2^2;
e3 = 1 - la3^2;
e4 = 1 - la4^2;
e5 = 1 - la5^2;
e6 = 1 - la6^2;
e7 = 1 - la7^2;
e8 = 1 - la8^2;
e9 = 1 - la9^2;
e10 = 1 - la10^2;
e11 = 1 - la11^2;
e12 = 1 - la12^2;
e13 = 1 - la13^2;
e14 = 1 - la14^2;
```

## *R Code to Compute the Weights for the $\bar{\chi}^2$ Distribution*

The weights for the $\bar{\chi}^2$ *distribution* were computed as the proportion of the number of correlations with $r > 1$ in each sample (i.e., the number of correlations that were larger than the boundary value in each sample) to the total number of simulated samples. To this end, the R code first read in the file "stoel_test.res," which contained the results from the Mplus simulation study. The number of correlations with $r > 1$ in each sample was then determined. In the present example, of the 10,000 samples, there were 703, 1322, 1960, 2152, 1807, 1383, and 673 cases with 0, 1, 2, 3, 4, 5, and 6 correlations, respectively, with $r > 1$. To compute the $\bar{\chi}^2$ distribution, we drew on the R code by Stoel et al. (2006). We therefore entered the weights .0703 : .1322 : .1960 : .2152 : .1807 : .1383 : .0673 in the "pmix" vector. These weights were used to combine the individual $\chi^2$ distributions with $df = 0$, $df = 1$, $df = 2$, $df = 3$, $df = 4$, $df = 5$, and $df = 6$ to compute the $\bar{\chi}^2$ distribution.

```
# This part of the syntax reads in the results file produced by Mplus as a data.frame

# specify the folder where the results file is located
# Note that R uses slashes "/" rather than back slashes "\" to separate folders
# Further, R is case sensitive

setwd("D:/PUBLIKATIONEN/2009_Reliability/analysen/WAIS-III_Spain/stoel_test/")

#enter the file name between the quotes
```

```
file.name<-"stoel_test.res"

#---------------------------------------------------
#This part of the syntax was written by Ulrich Keller (University of Luxembourg)

## Read text file
txt <- readLines(file.name)
## Remove leading spaces
txt <- sub("^ +", "", txt)
## Determine which lines contain iteration numbers
itnum.idx <- grep("^[0-9]+$", txt)
## Add imaginary last iteration number to end (needed later)
itnum.idx <- c(itnum.idx, length(txt) + 1)

## Iterate over iteration numbers (except imaginary last one)
num <- lapply(1:(length(itnum.idx) - 1), function(i) {
  ## Extract corresponding text lines
  txt.i <- txt[(itnum.idx[i] + 1):(itnum.idx[i + 1] - 1)]
  ## Convert to numeric and put together
  as.numeric(unlist(strsplit(txt.i, " +")))
})

## "num" is a list right now, so make a matrix
results <- data.frame(do.call(rbind, num))
#---------------------------------------------------




# To facilitate the analyses of the results file, reasonable variable names are provided
# according to the results saving information provided by Mplus

#RESULTS SAVING INFORMATION

#  Order of data
#    Parameter estimates
#     (saved in order shown in Technical 1 output)
#    Standard errors
#     (saved in order shown in Technical 1 output)
#    Chi-square : Value
#    Chi-square : Degrees of Freedom
#    Chi-square : P-Value
#    H0 Loglikelihood
#    H1 Loglikelihood
#    Number of Free Parameters
#    Akaike (AIC)
#    Bayesian (BIC)
#    Sample-Size Adjusted BIC
#    RMSEA : Estimate
#    SRMR


# the model that we investigated contains 62 parameters (p) and 62 corresponding standard
# errors (se; see Technical 1 output of the simulation study in Mplus)
parameters<-paste("p",1:62,sep="")
standard_errors<-paste("se",1:62,sep="")

# This section assigns variable names to the results file according to the
# results saving information provided by Mplus
names(results)<-
c(parameters,standard_errors,"chi2","chi2_df","chi2_p","H0","H1","free_par","AIC","BIC","adjBI
C","RMSEA","SRMR")

#### Determining the Weights for the Mixture Distribution

# parameters 57 to 62 contain the intercorrelations among
# the latent variables VC, PO, WM, and PS
# (see Technical 1 output of the simulation study in Mplus)
# To identify the number of boundary parameters, this part of the syntax analyzes whether
# a certain correlation is smaller or larger than 1.

for (i in 57:62) {
                      checkvar<-paste("p",i, sep="")
                      newvar<-paste("b",i, sep="")
                      results[,newvar] <- ifelse(results[,checkvar] < 1, c(0), c(1))
                     }
```

```
# compute the number of boundary parameters
# within each sample (i.e., for each row in the results file)

results$boundary<-rowSums(with(results,cbind(b57,b58,b59,b60,b61,b62)))

# Determining the weights. The results of this table are then to be entered
# as weights to compute the chi2-bar distribution
prop.table(table(results$boundary))

#The following syntax is adapted from Stoel et al., 2006
#retrieved on March 9, 2011, from http://dx.doi.org/10.1037/1082-989X.11.4.439.supp
#R-script for computing critical values of the distribution and nominal p-values of the
critical values of the standard distribution

#First run the following lines with the setup of the program:
## ---------------------------------------------------------------------------
#
qmchisq=function(thresh,nc=nc,df=df,pmix=pmix,pcrit=pcrit) {
prob=0
if (df[1]==0) {prob=pmix[1]
for (ic in 2:nc) {prob=prob+pmix[ic]*(pchisq(thresh,df[ic]))}    }
if (df[1]>0) {
for (ic in 1:nc) {prob=prob+pmix[ic]*(pchisq(thresh,df[ic]))}     }
f=(prob-(1-pcrit))^2
                                                          }
pmchisq=function(thresh,nc=nc,df=df,pmix=pmix) {
prob=0
if (df[1]==0) {
prob=pmix[1]
for (ic in 2:nc) {prob=prob+pmix[ic]*(pchisq(thresh,df[ic]))}
               }
if (df[1]>0) {
for (ic in 1:nc) {prob=prob+pmix[ic]*(pchisq(thresh,df[ic]))}              }
prob                                                            }
## ---------------------------------------------------------------------------


# the chi2_bar distribution is composed of seven chi2 distributions with df = 0 to df = 6
# this constitutes 7 components

# number of components = 7;
# these components enter the computation of the chi2_bar distribution

nc=7

# df=0, df=1, df=2, df=3, df=4, df=5, and df=6;
df=c(0,1,2,3,4,5,6)

# the weights used to combine the various chi2-distributions are:
# 0.0703 : 0.1322 : 0.1960 : 0.2152 : 0.1807 : 0.1383 : 0.0673

pmix=c(0.0703, 0.1322, 0.1960, 0.2152 ,0.1807, 0.1383, 0.0673)


pcrit=.05                # the significance level=.05

#
stheta=0
for (i in 1:nc) {stheta=stheta+pmix[i]*df[i]}
resth=optim(stheta,qmchisq,method="BFGS",nc=nc,df=df,pmix=pmix,pcrit=pcrit)
thresh=resth$par[1]          # provides critical value given alpha=.05
cat(' threshold given ',pcrit,' = ',thresh, "\n")
thresh=qchisq(.95,6)                         # nominal p-value of the critical value     for a
chi2 distribution with df = 6
thresh                       # nominal p-value of the critical value     for a chi2
distribution with df = 6
prob=1-pmchisq(thresh,nc=nc,df=df,pmix=pmix) # value of standard chi2(6) distribution at
alpha=.05
cat(' pcrit given ',thresh,' = ',prob, "\n")
prob=1-pmchisq(1408,nc=nc,df=df,pmix=pmix) # p value of the chi2-differences with respect to
the chi-bar-square distribution
cat(' pcrit given delta chi2 = 1408 = ',prob, "\n")
#
```

## References

Colom, R., Abad, F. J., Garcia, L. F., & Juan-Espinosa, M. (2002). Education, Wechsler's Full Scale IQ, and g. *Intelligence*, *30*, 449-462.

Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus User's Guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

R Development Core Team. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Stoel, R. D., Garre, F. G., Dolan, C., & van den Wittenboer, G. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods, 11*, 439-455.