2.1.1 Three Regular EFA approaches. H&S example

Regular EFA Using Three Mplus Approaches: Analysis of Two Examples

- TYPE = EFA
- ESEM (*1)
  - Asparouhov & Muthén (2009). Exploratory structural equation modeling. Structural Equation Modeling. 16, 397-438.
- PSEM with GEOMIN priors
  - Asparouhov & Muthén (2024). Penalized structural equation models. Structural Equation Modeling. 31, 429-454.
  - Asparouhov & Muthén (2025). Methodological advances with penalized structural equation models. Structural Equation Modeling. 32, 688-716.

We now turn to two examples. Slide 8 shows 3 Mplus approaches to EFA that we will use.

TYPE = EFA allows the analysis of a range of factors in one analysis, computing chi-square difference tests between number of factors.

The ESEM approach gives an equivalent approach for a specific number of factors and also serves as a bridge towards more general EFA settings to be discussed later. The 2009 article is the seminal source of the theory for ESEM.

The PSEM approach gives another equivalent approach using the GEOMIN priors that are the default for the TYPE = EFA approach. As shown in the two articles, this also serves as a bridge towards more general EFA settings to be discussed later.

Holzinger & Swineford (1937, 1939) Four Domains Measured by 19 Tests: Factor Loading Pattern

Test	Spatial	Verbal	Speed	Memory
Visual perception	X	0	0	0
Cubes	X	0	0	0
Paper form board	X	0	0	0
Flags	X	0	0	0
General information	0	X	0	0
Paragraph comprehension	0	X	0	0
Sentence completion	0	X	0	0
Word classification	0	X	0	0
Word meaning	0	X	0	0
Addition	0	0	X	0
Code	0	0	X	0
Counting groups of dots	0	0	X	0
Straight and curved capitals	0	0	X	0
Word recognition	0	0	0	X
Number recognition	0	0	0	X
Figure recognition	0	0	0	X
Object–number	0	0	0	X
Number–figure	0	0	0	X
Figure–word	0	0	0	X

On slide 9 we have our first example. This is the classic dataset used in the 1937 and 1939 Holzinger-Swineford papers given in the references. The figure shows a loading matrix for 19 tests with X marking hypothesized large loadings and 0 elsewhere. Four factors are measured by these tests: Spatial, Verbal, Speed, and Memory. Later on we will add 5 more tests and look at the bi-factor model of primary interest to Holzinger and Swineford.

In terms of analysis stages discussed on slide 4, having this hypothesized loading pattern corresponds to the late analysis stage of CFA but here we will primarily focus on EFA approaches. The dataset is available on the web paper page.

H&S Model with 4-factor CFA Using MLR

- Seventh- and eighth-grade students from two schools
- Grant–White (N=145): students from homes where the parents were mostly American born
- Pasteur (N = 156): students largely from working-class parents of whom many were foreign born and used their native language at home

Grant-White School

Model	Par's	LL	BIC	χ^2	Df	P	CFI
CFA	63	-9050	18412	217	146	0.000	0.927
Pasteur School							
CFA	63	-9909	20137	273	146	0.000	0.871

On slide 10, we will first show results from a 4-factor CFA using the loading pattern of the previous slide.

The dataset consists of 7th and 8th grade students from two schools, Grant-White and Pasteur. Each has a sample of about 150 students. As the table shows, both schools get a poorly fitting model as judged by chi-square with zero p-value. Given the moderate sample size, the poor fit cannot be ascribed to rejection due to a large sample. Also supporting the rejection, the MLR estimator is used which reduces the sensitivity of the chi-square test to nonnormality in the factor indicators.

CFA MLR Modification Indices for H&S Pasteur School

	M.I.	E.P.C.	Std E.P.C.	StdYX E.P.C.
BY Statements				
SPATIAL BY ADDITION	10.263	-0.301	-0.301	-0.301
SPATIAL BY FIGURER	13.969	0.344	0.344	0.344
VERBAL BY VISUAL	11.835	0.402	0.402	0.402
SPEED BY NUMBERR	12.957	-0.361	-0.361	-0.361

- Freeing the 4 cross-loadings still gives poor chi-square with zero p-value: Chi-2 (142) = 222
- A second round of freeing large modindices can be done but EFA may be a better alternative, giving results in one run
- See arguments in favor of EFA in the ESEM paper Asparouhov & Muthén (2009)

Slide 11 shows the key modification indices for the CFA run corresponding to unexpected cross loadings. The expected parameter change values for these are rather large in the standardized scale so these are substantively important misspecifications.

As the first bullet says, however, freeing these 4 cross loadings still results in poor fit. Instead of going further as guided by modification indices, EFA is a better approach in line with arguments mentioned on slide 4 in the ESEM paper quote of Brown.

TYPE = EFA Input for H&S

TITLE: Raw data from Holzinger-Swineford's monograph Grades VII and VIII of the Pasteur elementary school (n=156)
followed by Grant-White (n=145).
Gender, grade, and age information, and the 24 tests (Pasteur does not have tests 25 and 26)
Sources: Holzinger, K. J. & Swineford, F. (1939). A study in Factor
Analysis. The Stability of a Bi-Factor Solution. Supplementary Educational Monographs. Chicago, Ill.: The
University of Chicago.
Harman, H.H. (1976). Modern Factor Analysis. Third Edition. Chicago: The University of Chicago Press

DATA: FILE IS H-S Combined.txt;

VARIABLE: NAMES = names are id female grade agey agem school
visual cubes paper flags general
paragrap sentence wordc wordm addition code counting straight wordr figurer object numberf
figurew deduct numeric
problemr series arithmet;
USEVARIABLES =
visual-figurew;
USEOBSERVATIONS =
school eq 0; !school = 0 for
! Grant-White, 1 for Pasteur

ANALYSIS: **TYPE = EFA 3 5**;
! GEOMIN is the default rotation
ESTIMATOR = MLR;
! STARTS = 20;

PLOT: TYPE = PLOT3;

Bengt Muthén Exploratory Factor Analysis 12/ 139

Slide 12 shows the input for EFA where in the ANALYSIS command, EFA with 3, 4, and 5 factors is requested. GEOMIN is the default rotation.

For some special situations, it may be appropriate to add several random starting values using the STARTS option. Here, 20 such sets of starting values are used.

ESEM Input for H&S

ANALYSIS: ESTIMATOR = MLR;

MODEL: **f1-f4 BY visual-figurew(*1)**;

- ESEM is activated by (*1), which designates an EFA "block" of items that are indicators of the set of factors listed.
  The "1" is arbitrary and can be replaced with e.g. "a" or "efa"
- The ESEM approach to EFA allows the usual choices of rotations
- ESEM allows more general modeling including residual covariances and including other variables such as covariates
- ESEM generalizes to SEM settings where there can be more than one EFA block and where EFA blocks can be combined with CFA blocks

Slide 13 shows the ESEM approach for 4 factors. ESEM is activated by (*1), which designates an EFA “block” of items that are indicators of the set of factors listed, such as f1-f4 BY visual-figurew. The “1” is arbitrary and can be replaced with e.g. “a” or “efa”.

The ESEM approach to EFA allows the usual choices of rotations: orthogonal or oblique as well as bifactor.

The ESEM specification allows more general modeling including residual covariances and including other variables such as covariates. In fact, ESEM generalizes to SEM settings where there can be more than one EFA block and where EFA blocks can be combined with CFA blocks.

PSEM with GEOMIN Priors Input for H&S

```
ANALYSIS:       ESTIMATOR = MLR;
                ! some of the below settings might
                ! be needed:
                ! ITERATIONS = 10000;
                ! CONVERGENCE = 0.000001;
                ! STARTS = 50;

MODEL:          ! label the 4*19=76 factor loadings
                ! for which the GEOMIN rotation
                ! should be applied:
                f1-f4 BY visual-figurew*(a1-a76);
                f1-f4@1;

MODEL PRIORS:   a1-a76 ~ GEOMIN(4,1.0); ! GEOMIN settings
                ! are shown at the end of the Theory section
```

- The GEOMIN priors are applied to all factor loadings which gives the same rotated solution as the default for TYPE = EFA with 4 factors
- It is recommended to standardize the factor indicators before analysis so that the priors work optimally. This can be done using the STANDARDIZE = visual-figurew; option of the DEFINE command.

Slide 14 shows the input for PSEM with GEOMIN priors for the H&S data. This is our first PSEM example in that it uses priors. The commented lines explain the specifications.

In the MODEL command, parameter labels are given to all 76 factor loadings.

The MODEL PRIORS command specifies GEOMIN priors for the loadings. This gives essentially the same rotated solution as the default GEOMIN rotation for TYPE = EFA with 4 factors.

The GEOMIN setting "4" refers to the number of factors and 1.0 refers to the prior variance or rather 1/the penalty weight. GEOMIN settings are given in the EFA Theory section.

The factor indicators in the Holzinger-Swineford data have very different variances. We recommend standardizing the factor indicators before analysis so that the priors work optimally. This can be done using the STANDARDIZE option in the DEFINE command. For EFA, the standardization does not affect the chi-square testing.

As we will see, the GEOMIN priors approach is very convenient in that it can be used for other applications such as second-order analysis, bi-factor analysis, EFA in structural equation modeling settings, and EFA in longitudinal settings.

Summary of Model Fit Information using TYPE = EFA 3 5 for Grant-White with MLR

Model	# Parameters	Chi-Square	Df	P-Value
3-factor	92	207.070	117	0.0000
4-factor	108	116.071	101	0.1450
5-factor	123	98.474	86	0.1688

Models Compared	Chi-Square	Df	P-Value
3-factor against 4-factor	117.689	16	0.0000
4-factor against 5-factor	17.428	15	0.2939

- Choose the 4-factor model
- Chi-square difference testing with MLR uses the special approach described in `https://www.statmodel.com/chidiff.shtml`
- Asparouhov & Muthén (2024b) discusses alternative tests of the number of factors

Slide 15 shows the chi-square test of fit for 3, 4, and 5 factors. This is for the Grant-White school using MLR. Both a 4- and a 5-factor model fit well, but chi-square difference testing shows that 4 factors is sufficient in that it cannot be rejected when tested against 5 factors. So, we choose 4 factors.

With MLR, the chi-square difference testing uses the Satorra-Bentler approach that you find described on the Mplus website:
`https://www.statmodel.com/chidiff.shtml`

Alternative ways to test the number of factors in EFA is discussed in the Asparouhov & Muthén (2024) web note 25.

H&S Models with 4-factor CFA and EFA Using MLR

Grant-White School, N = 145

Model	Par's	LL	BIC	χ^2	Df	P	CFI	X-loads
CFA	63	-9050	18412	217	146	0.000	0.927	-
EFA	108	-8997	18532	116	101	0.145	0.985	5
CFA + 5	68	-9025	**18388**	167	141	0.068	0.974	

Pasteur School, N = 156

Model	Par's	LL	BIC	χ^2	Df	P	CFI	X-loads
CFA	63	-9909	20137	273	146	0.000	0.871	-
EFA	108	-9843	20232	136	101	0.011	0.964	8
CFA + 8	71	-9872	**20103**	193	138	0.001	0.944	

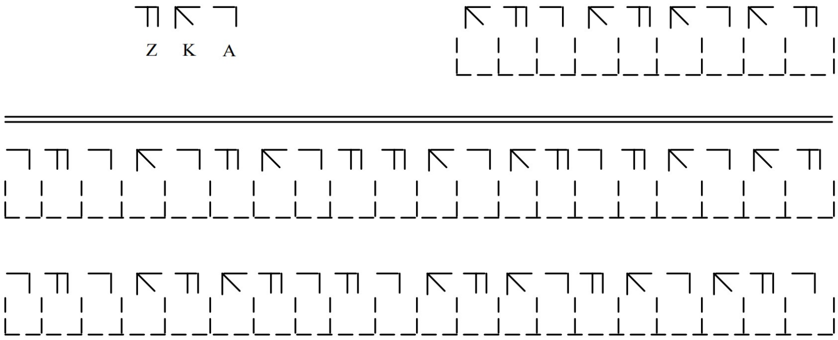
- X-loads refers to number of significant cross-loadings
- BIC = -2*LL + penalty, where penalty = # parameters * ln (N)

Slide 16 shows CFA and EFA test results for both schools. For both schools, EFA fits better in that its chi-square test gets a non-zero p-value unlike CFA. BIC, however, points to CFA. The BIC formula is shown at the bottom of the slide. This disagreement is not uncommon. In a sense, BIC shortchanges EFA. Not all EFA cross loadings are expected to be significant but these extra parameters are counted in the BIC penalty. As the table shows, EFA has a better log likelihood (LL) but many more parameters. The log likelihood improvement for EFA is not enough to compensate for the BIC penalty due to the large increase in number of parameters.

Out of the total of 57 cross loadings, Grant-White EFA has 5 that are significant and Pasteur has 8. Freeing these cross loadings in a modified CFA, the table shows that the best BIC is obtained for both schools. In this way, EFA provides an alternative to modification indices in terms of modifying a CFA. EFA may be closer to a well-fitting model than CFA so modifications of CFA based on EFA can be more likely to succeed than modifying the original CFA model by modification indices.

Examining Cross Loadings

- Code is an indicator of the Speed factor but also has a significant positive loading on the Memory factor for Grant-White
  - Requires matching letters to a set of figures



Bengt Muthén Exploratory Factor Analysis 17/ 139

Slide 17 shows an example of a cross loading in the analysis of H&S. This is one of the 5 significant cross loadings from the EFA of Grant-White. The indicator called code measures the speed factor but also loads on the memory factor. This makes sense in that the icons for the Z, K, A letters have to be remembered when quickly filling in the letters under the icons.

The cross loading is added to the CFA by saying: memory BY code;