

# METHODOLOGICAL ISSUES IN THE SEARCH FOR A FACTOR MODEL: EXPLORATION THROUGH CONFIRMATION

PIETER M. KROONENBERG  
 University of Leyden

CHARLES LEWIS  
 University of Groningen

**Key words:** *Confirmatory, cross-validation, EFAP, exploratory, factor analysis, Heywood case, LISREL*

**ABSTRACT.** The usefulness of some procedures suggested by Jöreskog for performing exploratory factor analysis is investigated through an in-depth analysis of some of the Holzinger-Swineford test data. In this case the procedures, which use both exploratory and confirmatory techniques, work not unreasonably, although they manage only just, and sometimes not quite, to steer away from misleading results.

In this paper we will comment on procedures to perform factor analysis that have been proposed in the literature by carrying out a detailed analysis on one particular data set. We will show that there are many pitfalls, surprises and uncertainties for the analyst on the way towards a solution, notwithstanding the seeming straightforwardness of the analysis in the literature. We will not give any description of techniques nor relate if, when, and why it is justified to take certain steps in factor analysis. For such information one should refer to standard texts on factor analysis such as Mulaik (1972), Harman (1967), Rummel (1970), and the paper by Frane and Hill (1976).

We restricted ourselves to the maximum likelihood method, given in Jöreskog (1966), and incorporated in the package of factor analysis routines EFAP (Jöreskog & Sörbom, 1976). Furthermore, we used the LISREL IV program (Jöreskog & Sörbom, 1978) for the confirmatory factor analysis for one or more groups.

One major advantage of using maximum likelihood over earlier approaches to factor analysis is that we obtain standard errors for our estimated loadings, factor correlations, and uniquenesses.

## Description of the Data

The data used in this study were originally described in a monograph by Holzinger and Swineford (1939). They collected scores on 26 psychological tests from seventh- and eighth-grade students in two Chicago-area schools: Pasteur and Grant-White.

Jöreskog (1969) selected nine of the original 26 tests (see Table I) and recomputed the correlation matrix on the total Grant-White sample, obtaining,

# METHODOLOGICAL ISSUES IN THE SEARCH FOR A FACTOR MODEL: EXPLORATION THROUGH CONFIRMATION

PIETER M. KROONENBERG  
 University of Leyden

CHARLES LEWIS  
 University of Groningen

**Key words:** *Confirmatory, cross-validation, EFAP, exploratory, factor analysis, Heywood case, LISREL*

**ABSTRACT.** The usefulness of some procedures suggested by Jöreskog for performing exploratory factor analysis is investigated through an in-depth analysis of some of the Holzinger-Swineford test data. In this case the procedures, which use both exploratory and confirmatory techniques, work not unreasonably, although they manage only just, and sometimes not quite, to steer away from misleading results.

In this paper we will comment on procedures to perform factor analysis that have been proposed in the literature by carrying out a detailed analysis on one particular data set. We will show that there are many pitfalls, surprises and uncertainties for the analyst on the way towards a solution, notwithstanding the seeming straightforwardness of the analysis in the literature. We will not give any description of techniques nor relate if, when, and why it is justified to take certain steps in factor analysis. For such information one should refer to standard texts on factor analysis such as Mulaik (1972), Harman (1967), Rummel (1970), and the paper by Frane and Hill (1976).

We restricted ourselves to the maximum likelihood method, given in Jöreskog (1966), and incorporated in the package of factor analysis routines EFAP (Jöreskog & Sörbom, 1976). Furthermore, we used the LISREL IV program (Jöreskog & Sörbom, 1978) for the confirmatory factor analysis for one or more groups.

One major advantage of using maximum likelihood over earlier approaches to factor analysis is that we obtain standard errors for our estimated loadings, factor correlations, and uniquenesses.

## Description of the Data

The data used in this study were originally described in a monograph by Holzinger and Swineford (1939). They collected scores on 26 psychological tests from seventh- and eighth-grade students in two Chicago-area schools: Pasteur and Grant-White.

Jöreskog (1969) selected nine of the original 26 tests (see Table I) and recomputed the correlation matrix on the total Grant-White sample, obtaining,

# METHODOLOGICAL ISSUES IN THE SEARCH FOR A FACTOR MODEL: EXPLORATION THROUGH CONFIRMATION

PIETER M. KROONENBERG  
 University of Leyden

CHARLES LEWIS  
 University of Groningen

**Key words:** *Confirmatory, cross-validation, EFAP, exploratory, factor analysis, Heywood case, LISREL*

**ABSTRACT.** The usefulness of some procedures suggested by Jöreskog for performing exploratory factor analysis is investigated through an in-depth analysis of some of the Holzinger-Swineford test data. In this case the procedures, which use both exploratory and confirmatory techniques, work not unreasonably, although they manage only just, and sometimes not quite, to steer away from misleading results.

In this paper we will comment on procedures to perform factor analysis that have been proposed in the literature by carrying out a detailed analysis on one particular data set. We will show that there are many pitfalls, surprises and uncertainties for the analyst on the way towards a solution, notwithstanding the seeming straightforwardness of the analysis in the literature. We will not give any description of techniques nor relate if, when, and why it is justified to take certain steps in factor analysis. For such information one should refer to standard texts on factor analysis such as Mulaik (1972), Harman (1967), Rummel (1970), and the paper by Frane and Hill (1976).

We restricted ourselves to the maximum likelihood method, given in Jöreskog (1966), and incorporated in the package of factor analysis routines EFAP (Jöreskog & Sörbom, 1976). Furthermore, we used the LISREL IV program (Jöreskog & Sörbom, 1978) for the confirmatory factor analysis for one or more groups.

One major advantage of using maximum likelihood over earlier approaches to factor analysis is that we obtain standard errors for our estimated loadings, factor correlations, and uniquenesses.

## Description of the Data

The data used in this study were originally described in a monograph by Holzinger and Swineford (1939). They collected scores on 26 psychological tests from seventh- and eighth-grade students in two Chicago-area schools: Pasteur and Grant-White.

Jöreskog (1969) selected nine of the original 26 tests (see Table I) and recomputed the correlation matrix on the total Grant-White sample, obtaining,



TABLE I  
Description of the Nine Tests

(1) Visual Perception	A nonlanguage, multiple-choice test.
(2) Cubes	A simplified test of spatial relations.
(3) Lozenges	Not described by Holzinger and Harman (1941); not included in their set of 24 variables. Presumably a spatial imagery test.
(4) Paragraph Comprehension	A silent reading test with comprehension measured by completion and multiple-choice questions.
(7) Sentence Completion	Multiple-choice test in which "correct" answers reflect good judgment.
(9) Word meaning	Multiple-choice vocabulary test.
(10) Addition	Speed of adding pairs of 1-digit numbers.
(12) Counting Groups of Dots	Four to seven dots arranged in patterns to be counted; test of perceptual speed.
(13) Straight and Curved Capitals	A series of capital letters with distinction required between those composed of straight lines only and those containing curved lines; test of perceptual speed.

*Note.* The descriptions are paraphrased from the appendix of Holzinger and Harman (1941), since we did not have the original monograph of Holzinger and Swineford (1939) at hand. The test names are given in the order they appear in all our analyses. The number to the left of each name is the number of the test in the original Holzinger and Swineford (1939) battery.

by the way, slightly different correlations from those of Holzinger and Swineford (largest absolute difference 0.023). These correlations are shown in Table II, along with the correlations from Jöreskog and Lawley (1968) who split the Grant-White sample into two subsamples ( $n_1 = 73$ ,  $n_2 = 72$ ), and recomputed the correlation matrices for the two subsamples.

Comparison of the correlations in the subsets with those in the complete set shows no extraordinary discrepancies, if one takes the 95 percent confidence intervals into account. The small sample size of the subsets nevertheless causes substantial differences in the values of the correlations across the subsets, especially for those near zero. The way the overall correlation matrix is organized makes it possible to do a rough visual factor analysis. By arranging the variables in three groups, that is,

group 1: visual perception, cubes, lozenges;

group 2: paragraph completion, sentence completion, and word meaning;

group 3: addition, counting dots, and straight/curved capitals,



TABLE II  
Correlation Matrices for the Holzinger-Swineford Data, Grant-White School

	GROUP I										GROUP II										GROUP III										
	VP	VP	VP	VP	VP	VP	VP	VP	VP	VP	PC	PC	PC	PC	PC	PC	PC	PC	PC	PC	WM	WM	WM	WM	WM	WM	WM	WM	WM	WM	SCC
Visual Perception (1)	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Cubes (2)	33	25	41	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Lozenges (3)	45	42	48	36	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Paragraph Comprehension (4)	34	28	23	22	43	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Sentence Completion (5)	31	26	16	13	29	30	72	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Word Meaning (6)	32	24	20	13	35	33	72	74	69	73	65	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Addition (7)	10	12	7	15	27	8	21	19	22	29	36	18	12	23	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Counting Dots (8)	31	25	17	18	24	15	10	2	22	28	20	12	12	30	60	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Straight-curved capitals (9)	49	58	25	15	37	31	31	38	26	40	25	27	24	29	42	53	46	63	100	100	100	100	100	100	100	100	100	100	100	100	100

Note. The correlation in each cell is based on: Complete set; Subset I; Subset I. All numbers have been multiplied by 100.

we see that each group more or less defines a factor. But the variables of each group have sizeable correlations with the other group variables.

### Basic Exploratory Analysis

One of the first questions one faces in factor analysis is how many factors underlie the data. Ideally, knowledge of the subject matter should give one the answer, at least approximately.

Conventionally, the question of the number of factors was decided by looking at the roots of the covariance (or correlation) matrix, and applying the skree "test," or by looking at the interpretability of the factors. The advent of maximum likelihood factor analysis has provided a stronger and more appropriate test for the purpose. This goodness of fit  $\chi^2$  test still is not perfect as it is sample size dependent, and is used for accepting null hypotheses rather than rejecting them. Fortunately, additional information is available in the EFAP output, such as the Tucker-Lewis reliability coefficient (Tucker & Lewis, 1973), and the matrix with so-called "residual correlations" (which are correlations among residuals).

For the data, a summary of the analysis with 1 through 4 factors is given in Table III. Inspecting this table, the analyses of the data seem to suggest two or three factors. If one would take Jöreskog & Lawley's (1968) general rule, that "it is probably wise to continue increasing the value of  $k$  (the number of factors) until the probability  $P$  exceeds 0.10" (p. 93), then a solution with three factors seems better than two. Also another rule, formulated in connection with confirmatory analysis, which advises to stop increasing the number of parameters (here: factors) if no significant improvement of the  $\chi^2$  is achieved by adding these parameters, suggests three factors. The decrease in the  $\chi^2$ -statistic from 62 to 9 is significant for 7 degrees of freedom, but the decrease from 9 to 3 (see Table III) is not.

On the other hand, the probability value of .67 at three factors gives us the uneasy feeling that we are overfitting with three factors. The larger-than-one value of the Tucker-Lewis coefficient points in the same direction. Finally, the guidelines from Harman (1967, p. 22) that the standard deviation of the residual correlations (here approximated by the average absolute residual correlation) should be about  $1/\sqrt{n}$  (here .08) makes one uncertain about three factors as well.

If it were possible, one would like to have something like  $2\frac{1}{2}$  factors. This may sound silly, but it is not, if we slightly rephrase the suggestion in terms of parameters to be estimated. With two factors, the solution is too constrained (i.e., there are not enough free parameters). The three-factor solution does not have enough constraints to fix the solution properly. In other words, there are so many parameters that even some idiosyncrasies of the data are taken into account in the estimates. The obvious solution is to put more constraints on

TABLE III  
Holzinger-Swineford Data—Exploratory Analyses

number of factors	degrees of freedom	$\chi^2$	probability level ( $p$ )	Tucker-Lewis coefficient $\rho$	"residual correlations"			comments
					average absolute	minimum	maximum	
1	27	176	.00	.56	.16	-.23	.55	
2	19	62	.00	.82	.09	-.27	.32	
3	12	9	.67	1.02	.03	-.07	.11	
4	6	3	.86	1.05	.02	-.07	.04	variable 7 Heywood case

the three-factor solution (i.e., fix some of the otherwise-to-be-estimated parameters). We will attack the problem in two ways, both starting with an exploratory analysis and ending with one or more confirmatory analyses. The first approach follows Jöreskog & Lawley (1968) by using the time-honored, but too seldom employed, practice of splitting one's sample into two parts, exploring one half of the data for patterns, or otherwise useful hypotheses, and confirming these with the other half. The second approach follows a procedure suggested by Jöreskog (1978). We will then discuss the results of the two approaches.

#### Approach 1: Split the Data into Halves

Jöreskog & Lawley used subset I for the exploratory analysis and subset II for the confirmatory analysis (for the correlation matrices of the two subsets, see Table II). There is, however, no a priori reason why subset I should have been chosen to be the exploratory set and subset II the confirmatory one. In the sequel we have performed the analysis both ways.<sup>1</sup>

#### Exploration on the Subsets

The results of these analyses are summarized in Table IV A and B.

<sup>1</sup>It should be mentioned that Jöreskog & Lawley used the forerunners of the programs we used; that is, for exploration they used UMLFA (Jöreskog, 1966), now superseded by EFAP; for confirmation they used RMLFA (Jöreskog & Gruvaeus, 1967), now subsumed in LISREL IV.



TABLE IV A  
 Subset I (Exploration Sample in Jöreskog & Lawley)—Exploratory Analysis<sup>a</sup>

number of factors	degrees of freedom	$\chi^2$	probability level ( $p$ )	Tucker-Lewis coefficient $\rho$	residual correlation			comments
					average absolute	minimum	maximum	
1	27	101	.00	.54	.17	-.32	.54	
2	19	46	.00	.76	.11	-.29	.38	$\psi_8 = 0$ (near Heywood case)
3	12	5	.95	1.10	.04	-.10	.09	
4	6	2	.88	1.10	.02	-.06	.10	

<sup>a</sup>Identical to solutions of Jöreskog and Lawley (1968, p. 93, 94).

#### Subset I.

- the probability,  $p$ , exceeds .10 at three factors;
- the decrease  $\chi^2$  is significant at the 5 percent level in going from two to three factors, but not from three to four factors;
- the probability level at three factors (.95) is uncomfortably high, indicating too good a fit; at two factors the .00 is not acceptable;
- the Tucker-Lewis coefficient is too low at two factors and too high at three;
- the mean absolute residual correlation is about right ( $1/\sqrt{n} = .12$ ) for two factors, and rather low for three factors.

Conclusion for subset I: Two to three factors, or  $2\frac{1}{2}$  if you like.

#### Subset II

- the probability,  $p$ , exceeds .10 at three factors;
- the decrease in  $\chi^2$  is significant at the 5 percent level in going from two to three factors, and is not significant in going from three to four factors;
- the probability level is reasonable at three factors, but not acceptable at two factors;
- the Tucker-Lewis coefficient is acceptable at two factors, but getting rather high at three factors;
- the mean absolute residual correlation is about right at two factors, and rather low for three factors.

TABLE IV B  
 Subset II (Confirmation Sample in Jöreskog & Lawley)—Exploratory Analysis

number of factors	degrees of freedom	$\chi^2$	probability level ( $p$ )	Tucker-Lewis coefficient $\rho$	residual correlation			comments
					average absolute	minimum	maximum	
1	27	106	.00	.56	.17	-.26	.59	
2	19	38	.01	.85	.10	-.25	.39	
3	12	15	.26	.97	.07	-.20	.23	
4	6	4	.71	1.06	.02	-.16	.08	2 Heywood cases variables 8, 9

Conclusion for subset II: Again two to three factors. Conclusion on the basis of the two subsets together: Use a three-factor model, but restrict the number of parameters to be estimated.<sup>2</sup>

#### Confirmation with the Other Subset

As mentioned before, Jöreskog & Lawley used subset I for their exploration, and subset II for confirmation. The result of the subset I varimax rotated solution is shown in Table VA, alongside the subsequent oblique rotations from UMLFA (VB) and the EFAP analysis (VC). Obviously, on the basis of both the varimax and oblique rotations, Jöreskog & Lawley decided to use *pattern b* (see Table VI) as their factor pattern to be tested with the confirmation sample (subset II). From line 5 of Table VII it can be seen that the results were entirely satisfactory from a fitting point of view ( $\chi^2 = 30$ ,  $df. = 23$ ,  $p = .15$ ). However, when we tried to replicate their results we ran into unexpected troubles. At the time of the first analysis, we did not have the results of Jöreskog & Lawley at hand, and so we based our confirmatory analysis on factor *pattern c* (see Table VI) provided by the promax rotation of

<sup>2</sup>Note, by the way, the sometimes large differences in the results between the two subsets, especially in the  $\chi^2$  fit of the three-factor model, and in the unique variances of variables 8 and 9 in the four-factor model (subset I:  $\hat{\psi}_8 = .2$  and  $\hat{\psi}_9 = .4$ ; subset II:  $\hat{\psi}_8 = \hat{\psi}_9 = .0$ ;  $\hat{\psi}_i$  indicates the estimated unique variance of variable  $i$ ).

TABLE V  
Holzinger-Swineford Data—Exploratory Factor Patterns

Subset I Variable	Orthogonal (Varimax) UMLFA & EFAP (A)			Oblique UMLFA (B)			Oblique (Promax) EFAP (C)		
	f1	f2	f3	f1	f2	f3	f1	f2	f3
1	.61	.30	.23	.60	.14	.14	.58	.15	.14
2	.60	.12	.06	.63	-.02	.00	.62	-.01	-.01
3	.72	.18	-.02	.74	.03	.11	.75	.04	-.12
4	.17	.82	.11	-.02	.88	-.06	.02	.86	-.06
5	.18	.75	.21	.01	.77	.05	.03	.75	.06
6	.26	.76	.16	.10	.78	.00	.13	.76	.00
7	-.22	.22	.76	-.24	.14	.78	-.34	.13	.80
8	.21	.12	.82	.26	-.09	.82	.14	-.08	.84
9	.40	.14	.65	.44	-.08	.63	.35	-.06	.64

TABLE VI  
Patterns for the Factor Loading Matrix ( $\Lambda$ )  
Pattern a: Independent clusters

Variable	f1	f2	f3	
1	x	0	0	0 = zero loading x = loading to be estimated
2	x	0	0	
3	x	0	0	
4	0	x	0	pattern b: $\lambda(9, 1)$ is also estimated
5	0	x	0	
6	0	x	0	
7	0	0	x	pattern c: $\lambda(9, 1)$ and $\lambda(7, 1)$ are also estimated
8	0	0	x	
9	0	0	x	



TABLE VII  
LISREL IV Analysis of Subsets

## A. Subset I

	number of factors	degrees of freedom	$\chi^2$	probability level ( $p$ )	number of nonsignificant loadings	type of factor pattern	comments
1.	3	23	26	.31	0	<i>b</i>	pattern suggested by subset II exploration
2.	3	22	17	.75	1	<i>c</i>	pattern suggested by its own exploration

## B. Subset II

3.	3	22	29	.14	4	<i>c</i>	pattern suggested by subset I exploration, $\hat{\psi}_8 = -.93!!$
4.	3	23	29	.17	2	<i>c</i>	$\psi_8$ fixed at zero
5.	3	23	30	.15	1	<i>b</i>	pattern suggested by its own exploration and Jöreskog and Lawley

## C. Subset I &amp; Subset II

6.	3	68	72	.33	0	<i>b</i>	common pattern of subsets I and II
----	---	----	----	-----	---	----------	------------------------------------

EFAP, using the same considerations as Jöreskog & Lawley, viz., as the sample size is rather small the sampling variability will be large; therefore only factor loadings larger than, say .30, should be taken into account. Our confirmatory analysis (see line 3 of Table VII) was far from satisfactory as it came up with a *negative estimate* ( $-.93!$ ) of the unique variance of variable 8. In other words, the solution was improper and none of the estimates could be trusted.

Technically, an improper solution suggests that the function to be minimized has no satisfactory value inside the admissible region (here the region where all unique variances are greater than or equal to zero). Hence, all one can do is search for a minimum value on the boundary, which implies setting the value of the deviant unique variance equal to zero. The estimates of the other values are then correctly computed. This was done, and the result is given in line 4 of Table VII. In general the existence of an improper solution or Heywood case

indicates that there is a conflict between the model (i.e., the hypothesized structure) and the data. The reason for the improper solution is not clear cut, but it is probably the following: inspection of the correlation matrix of subset II (Table II) shows that variable 8 has its highest correlations with variables 7 and 9, and its correlations with the group I variables are in between those of 7 and 9. Notwithstanding, 7 and 9 are "allowed" to have loadings on the group I factor, and 8 is not. It seems to us that the resulting conflict might have produced the improper solution. A thorough discussion of improper solutions can be found in Van Driel (1978).

The other, reverse, investigation, that is, exploring subset II, and confirming the resulting *pattern b* with subset I, ended peacefully with a  $\chi^2 = 26$  for 23 degrees of freedom ( $p = .31$ ) (line 1 of Table VII). As expected, *pattern c* fits even better ( $\chi^2 = 17$ ,  $df. = 22$ ,  $p = .75$ ), probably producing an overfit.

The conclusion of the two analyses is that *pattern b* is an acceptable one for both subsets. Another question is, does there exist a *model* which is compatible with both subsets? In other words, can we find a pattern matrix, uniqueness, and factor correlations which can be satisfactorially and simultaneously estimated from both subsets? Such a question can be answered using the so-called *simultaneous factor analysis* for several groups (Jöreskog, 1971). This analysis can also be performed with LISREL IV without any problems. The results there for our data were surprisingly well behaved ( $\chi^2 = 72$ ,  $df. = 68$ ,  $p = .33$ , no insignificant loadings), showing that there is no reason to suppose that the two subsets came from different populations, and, moreover, that the procedure followed in this section can lead to reasonable answers.

Nevertheless, the unfortunate results of one of the analyses shows the sensitivity of the procedure quite clearly. It is very difficult to pinpoint in this case what went wrong: if the rotational procedure is to blame, the small sample size, or something else. It is clear, in any case, that a blind following of procedures such as the above can lead to very nasty results. Setting aside the question of the sensitivity of the procedure, some further remarks are in order with respect to the procedure itself. There is no reason why one should stop with the simple *cross-validation* suggested by Jöreskog & Lawley. As Mosteller & Tukey (1977, p. 38) argue, this one-way procedure is inspired by hypothesis testing. However, in cases such as this, one is not really concerned with hypothesis testing but with estimation. One tries to find a reasonable model, and estimates of its parameters. Therefore, the least one can do, as we did, is interchange the two subsets, do the analyses again, and combine the information of the two subsets. Extending the procedure brings one automatically either to the "jackknife" (e.g., Mosteller & Tukey, 1977, p. 133 ff.) or to the "bootstrap" (Efron, 1979). The problem with the jackknife procedure in factor analyses such as this one is computer time. Small problems like the present one could be done with a few groups, but with any sizeable set of variables the computing costs might easily get out of hand.

TABLE VIII  
 LISREL IV Analysis of Complete Set, Based on Analyses of Subsets  
 Estimated Loadings, Uniquenesses, and Factor Intercorrelations ( $\text{all} \times 100$ ),  
 Together with their Standard Errors

		Factors			
	Var	f1	f2	f3	unique
1.	VP	71(09)			49(09)
2.	C	49(09)			76(10)
3.	L	66(09)			56(09)
4.	PC		87(07)		25(05)
5.	SC		83(07)		31(05)
6.	WM		82(07)		32(05)
7.	A			68(09)	53(09)
8.	CD			86(09)	26(11)
9.	SCC	45(09)		44(09)	45(07)

factor correlations

2 55(08)

3 38(10) 24(10)

$$\chi^2_{23} = 28.1, p = .21$$

Final Estimates

Having found an acceptable pattern, the only question that remains is, how to obtain the final estimates. Do we take them from the single subsets, the two subsets simultaneously, or from the complete set? In our eyes the only acceptable estimates are those from the complete set; after all, that was the set of data for which we wanted to have estimates. These estimates, together with their standard errors, are given in Table VIII.

Approach 2: Exploration Through Confirmatory Analysis

An alternative to splitting the sample, fitting on one half and testing on the other half, is to test several models sequentially on the complete sample and adopt a model that fits reasonably well. Since there is no independent check on the result (as there was in the split half approach), extra care must be taken if one is to avoid capitalizing on characteristics of the sample which are actually due only to chance. In particular, one should not feel free to test every possible model. Instead, one should follow a reasonable set of guidelines in selecting



models for testing. Here we follow the suggestions of Jöreskog which were referred to at the end of the Basic Exploratory Analysis section.<sup>3</sup>

### Steps in Testing

- (1) *Find an unrestricted solution.* Perform an exploratory ML factor analysis, plus varimax rotation, using EFAP. The resulting matrix of loadings plus the uniquenesses for the complete data set is given (Table IXA).
- (2) *Define a rotation to be used in LISREL.* Find the largest loading in each column of the varimax matrix (underlined in Table IXA), set to zero the other loadings in the same rows, and perform a LISREL analysis. The result (Table IXB) is an oblique (*reference variable*) rotation of the varimax solution with no theoretical change in the fit of the model. (The actual discrepancy between the two  $\chi^2$  values is small and presumably reflects differences in algorithms, together with limited numerical precision.)
- (3) *Make small loadings zero to simplify model.* Examine the "t-statistics" provided for the loadings by LISREL (and given in Table IXC) and set to zero all loadings which do not exceed 2.58 in absolute value.<sup>4</sup> (These are underlined.) Perform a new LISREL analysis with these additional restrictions. (The result is given in Table IXD.) The difference in  $\chi^2$  between this and the previous analysis should not exceed the upper 5 percent point of the  $\chi^2$  distribution with degrees of freedom equal to the number of new restrictions introduced. (In the present case, 10 restrictions were introduced and the 5 percent point of  $\chi^2_{10}$  is 18.31. The actual difference is  $25.6 - 9.7 = 15.9$ , so there is no difficulty with accepting the more restricted model.)
- (4) *Examine the appropriateness of the simplified model.* Look at the first derivatives of the log likelihood with respect to all the loadings which have been fixed at zero in the previous LISREL analysis. (The derivatives are given in Table IXE.) Drop the restrictions on the loading whose derivative is largest in absolute value (underlined in Table IXE), and perform another LISREL analysis. (The results are given in Table IXF for our case.)

<sup>3</sup>Jöreskog's (1978) published guidelines differ slightly from those followed here, which are actually based on an earlier verbal discussion given by Jöreskog at a seminar in Groningen in 1978. The differences are noted in the subsequent discussions as they arise.

<sup>4</sup>Note: This is the two-tailed .01 value from a standard normal distribution, not from a t-distribution. Asymptotically, the "t-statistics" quoted by LISREL should follow a normal distribution. Nonasymptotically, there is no theory that suggests they should be t-distributed, so reference to the normal distribution seems to be the only choice.

TABLE IX  
Illustrations of Approach 2 (all loadings  $\times 100$ )

## (A) EFAP ML, varimax rotation

var.	f1	f2	f3	$\hat{\psi}$
1	<u>66</u>	20	16	50
2	50	11	05	74
3	64	21	08	54
4	23	<u>84</u>	08	24
5	18	79	19	30
6	23	79	07	32
7	-04	18	76	39
8	27	-00	<u>78</u>	32
9	48	19	52	46

$df = 12, \chi^2 = 9.3, p = .67$

## (B) Reference variable solution

var.	f1	f2	f3	$\hat{\psi}$
1	71	00	00	50
2	58	-06	-08	74
3	72	01	-10	54
4	00	87	00	24
5	-10	86	14	30
6	02	82	-01	32
7	-44	32	88	39
8	00	00	83	32
9	32	11	47	46

$df = 12, \chi^2 = 9.7, p = .64$

## (C) "t-statistics" of reference variable solution

var.	f1	f2	f3
1	8.13	-	-
2	3.76	<u>-48</u>	<u>-68</u>
3	4.51	<u>.06</u>	<u>-75</u>
4	-	12.41	-
5	<u>-92</u>	8.99	<u>1.66</u>
6	<u>.15</u>	8.94	<u>-09</u>
7	<u>-2.53</u>	-2.62	5.87
8	-	-	9.60
9	2.68	<u>1.10</u>	4.90

## (D) (More) restricted solution

var.	f1	f2	f3	$\hat{\psi}$
1	71	00	00	50
2	49	00	00	76
3	66	00	00	57
4	00	87	00	24
5	00	83	00	31
6	00	82	00	32
7	00	13	62	57
8	00	00	92	16
9	47	00	39	47

$df = 22, \chi^2 = 25.6, p = .27$

(continued on next pg.)

TABLE IX (Part II)

## (E) Derivatives of restricted solution

var.	f1	f2	f3
1	.00	.03	-.06
2	.00	.04	.03
3	.00	-.02	.04
4	-.02	.00	.12
5	-.00	.00	-.19
6	-.01	.00	.06
7	.15	.00	.00
8	-.10	.03	.00
9	.00	-.06	.00

## (F) Less restricted solution

var.	f1	f2	f3	$\hat{\psi}$
1	71	00	00	50
2	49	00	00	76
3	66	00	00	57
4	00	87	00	24
5	00	81	10	31
6	00	82	00	32
7	00	13	63	56
8	00	00	91	18
9	47	00	40	47

$$df = 21, \chi^2 = 22.6, p = .36$$

(G)  $\chi^2$  table

Source	df	$\chi^2$
More restricted (Table IXD)	22	25.6
Less restricted (Table IXF)	21	22.6
difference	1	3.0

$$\chi^2_{1,.05} = 3.8$$

(5a) *If simpler model is satisfactory, stop.* If the difference in  $\chi^2$  between the last two analyses does not exceed the 5 percent level of the  $\chi^2$  distribution with one degree of freedom (3.8), there is no reason to reject the hypothesis that the previously fixed loading is zero. Therefore, the more restricted (i.e., simpler) model is acceptable and analysis should stop. (See Table IXG for results in our case.)

(5b) *If simpler model is too restrictive, relax it.* If the difference  $\chi^2$  in step (5a) does exceed 3.8, repeat step 4 with the most recent LISREL results (here this would have been the solution of Table IXF) to see if the current model should be relaxed still further (i.e., if another restriction should be dropped).



## Discussion of Steps

- (1) The first step, finding an unrestricted solution, provides a basis for all subsequent LISREL analyses. It provides information on the number of factors and the rotation to be used. From this point of view, the subsequent steps could be seen as providing a refinement to the usual exploratory approach, as it was applied to the exploratory subsets above.
- (2) The second step, defining a rotation, is a bit artificial, but provides a way of getting the exploratory analysis into LISREL with relatively little distortion. It could be questioned whether the varimax rotation is the best one to take as a starting point.<sup>5</sup> In our case, the promax rotation provided by EFAP suggested the use of variables 1, 4 and 7 as reference variables instead of the 1, 4 and 8 actually used. When the subsequent steps were followed, but now on the promax reference solution, the final result given in Table XA was obtained. (The earlier result is given as well for comparison purposes.) The most striking difference is in the  $\chi^2$  values, with the value based on promax looking almost too good to be true. Nonetheless, it is worth remarking that all loadings in the "promax" solution of Table IX are significant at the .01 level, and that trying to fix one or more of them to zero results in a significant  $\chi^2$  difference. This is not to suggest that we find the loadings based on promax a more attractive final solution than those based on the varimax. We do not. (Interpretation is more difficult.) Instead, the point is that there is apt to be more than one "acceptable" model in any given situation, and that the guidelines which one adopts in one's search can make a real difference in which of these models one ends up with.
- (3) The third step, examining the "t-statistics" for the loadings, is designed to produce a simpler pattern of loadings without increasing the resulting  $\chi^2$ -value beyond what might be expected by chance (at the .05 level). Like the previous step, however, it has an element of artificiality, namely the choice of the .01 level for checking the individual statistics.<sup>6</sup> In our case, the loadings  $\lambda_{71}$ ,  $\lambda_{72}$ , and  $\lambda_{91}$  all have "t-statistics" close to the critical value of  $\pm 2.58$ . If the critical value were changed to  $\pm 2.70$ , all three loadings would be set to zero; if  $\pm 2.65$  were used, only  $\lambda_{71}$  and  $\lambda_{72}$  would be set to zero; with  $\pm 2.58$ , only  $\lambda_{71}$  was set to zero; finally, if  $\pm 2.50$  were

<sup>5</sup>Jöreskog (1978) suggests promax, but actually uses varimax in an illustration.

<sup>6</sup>Jöreskog (1978) suggests .05. This would give the same result as the critical value of  $\pm 2.50$  in the above, and makes no difference for the promax. In our view, .01 seems more reasonable, however, as the "t-statistics" are considered separately, instead of simultaneously.

TABLE X  
Solutions Starting From Different Rotations, Loadings  $\times 10$

A. From promax, variables 1, 4 and 7 used for reference

	1	2	3
1	7		
2	5		
3	7		
4		9	
5		8	
6		8	
7			8
8	4	-3	7
9	5		4

$$\chi^2_{21} = 15, p = .84$$

B. From varimax, variables 1, 4 and 8 used for reference

	1	2	3
1	7		
2	5		
3	7		
4		9	
5		8	
6		8	
7		<u>1</u>	6
8			9
9	5		4

$$\chi^2_{22} = 26, p = .27$$

Note. Differences for loadings on variables 7 and 8 have virtually no effect on the loadings for the remaining variables.

Those free in one solution but fixed in the other are underlined.

used none of the three loadings would be set to zero. The solutions corresponding to these four possibilities appear in Table XI. The difference between the  $\chi^2$  for solution XIA and that for the unrestricted case is 41 and has 12 degrees of freedom. Clearly, we have gone too far in using a critical value of 2.70. If, nonetheless, we follow the procedure of steps 4 and 5 for this solution, we find that  $\lambda_{91}$  has the largest derivative, leading us to examine solution XIB. With a  $\chi^2$  difference of 23 and one degree of freedom, we must reject the null hypothesis (per step 5b) that  $\lambda_{91}$  is zero. When steps 4 and 5 are repeated for solution B, no further improvement occurs. Thus XIB represents a satisfactory solution. It should be noted that XIB was also the solution arrived at via the split-half method. Solution XIC is our standard result, and has been included for reference purposes. Note that it does *not* represent a significant improvement over B ( $\chi^2$  difference of two with one degree of freedom), so the null hypothesis that  $\lambda_{72} = 0$  cannot be rejected. Finally, consider solution XID. Its  $\chi^2$  difference from the unrestricted solution is only five, with nine degrees of freedom—too good to be true. Following steps 4 and 5, it remains an

TABLE XI  
Results of Varying Critical Value  
Loadings  $\times 10$  for Four Different Models

	A 2.70			B 2.65			C 2.58			D 2.50		
	1	2	3	1	2	3	1	2	3	1	2	3
1	7			7			7			7		
2	5			5			5			5		
3	7			7			7			7		
4		9			9			9			9	
5		8			8			8			8	
6		8			8			8			8	
7			7			7	<u>1</u>	6	<u>5</u>	<u>3</u>		9
8			8			9		9				8
9			7	<u>4</u>		<u>4</u>	<u>5</u>		<u>4</u>			<u>5</u>
<hr/>												
	$\chi^2_{24} = 51$ $p = .001$			$\chi^2_{23} = 28$ $p = .21$			$\chi^2_{22} = 26$ $p = .27$			$\chi^2_{21} = 15$ $p = .81$		

Note. Those free in one solution but fixed in another are underlined.

"acceptable" solution. In terms of goodness of fit and difficulty of interpretation, solution XID bears some resemblance to the solution based on the promax rotation, given in Table XA. The introduction of two more acceptable solutions via slight changes in the critical value for the "t-statistics" reinforces the points made in the discussion of step 2 concerning nonuniqueness of results and sensitivity to modifications.

- (4) The fourth and fifth steps, freeing one of the loadings which had previously been set to zero and checking the effect of this, have been included to make sure we have not gone too far in our simplification of the model (i.e., put in too many zeros). A large derivative with respect to a fixed loading suggests that a substantial decrease in the  $\chi^2$  may be obtained by freeing that loading.<sup>7</sup>

<sup>7</sup>Jöreskog (1978) suggests using .01 instead of .05 for the choice of a critical  $\chi^2$  value. This has no effect on the present analyses. Here we agree with Jöreskog about his .01. See footnote 6.



Two comments seem worthwhile here. First, it is also possible at this stage that we have not gone far enough with our simplification and that a repeat of step 3 (examining "t-statistics") would be worthwhile. In our case, this would take us from solution C to solution B of Table XI. All loadings in solution B have clearly significant "t-statistics," so the process would stop here.

A second comment is that freeing the loading with the largest derivative does not necessarily result in the greatest improvement in  $\chi^2$ . From Table IXE,  $\lambda_{71}$  has only the second largest derivative. When it is freed, however, the result is solution D of Table XI. The  $\chi^2$  difference between XIC and XID is 11 with one degree of freedom, a much greater difference than that obtained by freeing  $\lambda_{53}$  (the parameter with the largest derivative), and a change that is clearly significant. Thus, by the standards of step 5, we should reject the hypothesis that  $\lambda_{71}$  is zero, and adopt solution XID.

#### *Final Comments on the Procedure and its Results*

In our opinion, solutions XID and XB represent over-fitting based on chance characteristics of the sample being studied. To put it another way, we would expect the fit of these two models to be *much* worse when applied to a new sample. One of the functions of any set of guidelines for exploratory factor analysis should be to steer us away from such solutions. It is to the credit of the Jöreskog guidelines that (in their unmodified form) they did *not* lead us to either of these solutions in the present case. We did, however, come uncomfortably close, and there is no assurance that we might not get into trouble another time. This is a point to keep in mind when using any inferential procedure. The best we can hope for is that a procedure will do well by us *most* of the time.

#### **Results of the Two Approaches and yet Another One**

We now briefly discuss and compare the results of the analyses performed in the "Approach I" and "Approach II" sections. The solutions (estimated on the complete data set) are given in detail in Table XIIA and B.

#### *Factor Loadings*

In comparing these results, the main point is that there is essentially no difference between them. The one extra loading in the "confirmatory" solution (that of the addition test on the second factor) is estimated at .13 with a standard error of .08. As noted in the previous section a slight modification of Jöreskog's guidelines would have resulted in fixing this parameter at zero, in which case we would indeed have adopted identical solutions via the two approaches. From an interpretational point of view, it definitely is preferable not to have an addition test loading on a factor whose largest loading comes from a paragraph comprehension test. Consequently, the "split-half" solution is the one which will be discussed further.

TABLE XII  
Three Complete Factor Solutions

A. From split-half procedure					B. From confirmatory procedure					C. From outside knowledge				
Var.	f1	f2	f3	unique	Var.	f1	f2	f3	unique	Var.	f1	f2	f3	unique
VP	71(09)			49(09)	VP	71(09)			50(09)	VP	73(09)			47(09)
C	49(09)			76(10)	C	49(09)			76(10)	C	49(09)			76(10)
L	66(09)			56(09)	L	66(09)			57(09)	L	67(09)			56(09)
PC		87(07)		25(05)	PC		87(07)		24(05)	PC		87(07)		25(05)
SC		83(07)		31(05)	SC		83(07)		31(05)	SC		83(07)		31(05)
WM		82(07)		32(05)	WM		82(07)		32(05)	WM		83(07)		32(05)
A			68(09)	53(09)	A		13(08)	62(09)	57(10)	A			78(10)	40(12)
CD			86(09)	26(11)	CD			92(10)	16(14)	CD	28(08)		71(09)	38(10)
SCC	45(09)		44(09)	45(07)	SCC	47(09)		39(09)	47(05)	SCC	56(08)		45(08)	44(07)
factor correlations					factor correlations					factor correlations				
2	55(08)				2	56(08)				2	52(08)			
3	38(10)	24(10)			3	39(10)	18(10)			3	—	—		
$\chi^2_3 = 28.1, p = .21$					$\chi^2_2 = 25.6, p = .27$					$\chi^2_4 = 28.4, p = .25$				

Note. Loadings, uniquenesses, and the correlations between the first and second factors are all virtually the same for all solutions.

As presumably intended by Holzinger & Swineford (1939), and as found by previous investigators, these nine variables determine what seem to be a *spatial relations* factor, a *verbal* factor, and a *speed* factor. In addition, the straight-curved capitals test (nominally a pure speed test) has a medium loading on the spatial relations factor. Since distinguishing straight from curved lines has a perceptual character to it, this loading does not seem to give us too much difficulty from an interpretational point of view. It might be argued that the same logic would suggest the possibility of a loading for the counting dots test on the spatial relations factor, but this does not seem to be a serious complaint (see also below).

#### *Factor Correlations*

With any oblique factor solution, there should be some discussion of the correlations among the factors. There seems to be nothing unusual in the present case. Positive correlation among abilities is common in unselected groups such as the school children who took these tests. One might have hoped the speed factor would have little relation to the other two, and indeed, the relevant correlations are not large (see also below).

#### *Uniquenesses*

The unique variances range from a negligible value (for counting dots) to a distressingly large one (for cubes). There is nothing wrong with small uniquenesses, at least when they are associated with highly reliable tests. (Holzinger & Harman, 1941, quote a reliability of .94 for counting dots.) In the same sense, large uniquenesses are not surprising when they are associated with unreliable tests. (The reliability for cubes is given as .57.<sup>8</sup>) In any event, the primary goal of factor analysis is to account for *intercorrelations* among variables—not their variances—and the “split-half” solution yields residual correlations which are no greater (and no smaller, as well) than might be expected by chance if it were the true model for the population. (This is what the nonsignificant  $\chi^2$  value tells us.)

#### *Yet Another Solution*

Before leaving the problem, it seems important to say that there is a place for outside knowledge, which can never be filled completely by exploratory techniques, no matter how refined. In the present case, it was noted that theoretical considerations might suggest a nonzero loading for counting dots on the spatial relations factor, and that perhaps the speed factor should be uncorrelated with the other two. Such modifications were introduced, and the result (complete with standard errors) appears in Table XIIC. It, again, provides a satisfactory solution, and is in many respects very similar to the other two solutions.

<sup>8</sup>It might be noted in passing that, of the first group variables, cubes has consistently the lowest correlations with all other variables (see Table II). In other words, it does not have very much variance in common with the other variables.



## References

- Efron, B. Bootstrap methods: Another look at the Jackknife. *Annals of Statistics*, 1979, 8, 1-26.
- Frane, J. W., & Hill, M. Factor analysis as a tool for data analysis. *Communications in Statistics (Part A)—Theory and Methods*, 1976, 5, 487-506.
- Harman, H. H. *Modern factor analysis* (2nd. ed.). Chicago: University of Chicago Press, 1967.
- Holzinger, K. J., & Harman, H. H. *Factor analysis*. Chacago: University of Chicago Press, 1941.
- Holzinger, K. J., & Swineford, F. A. *A study in factor analysis: The stability of a bi-factor solution*. (Supplementary Educational Monographs, No. 48). Chicago: University of Chiacgo, 1939.
- Jöreskog, K. C. UMLFA—A computer program for unrestricted maximum likelihood factor analysis. *Research memorandum*, 66-20. Princeton, N.J.: Educational Testing Service, 1966.
- Jöreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 1969, 34, 183-202.
- Jöreskog, K. G. Simultaneous factor analysis in several populations, *Psychometrika*, 1971, 36, 409-426.
- Jöreskog, K. G. Structural analysis of covariance and correlation matrices. *Psychometrika*, 1978, 43, 443-477.
- Jöreskog, K. G., & Gruvaeus, G. RMLFA—A computer program for restricted maximum likelihood factor analysis. *Research memorandum*, 67-21. Princeton, N.J.: Educational Testing Service, 1967.
- Jöreskog, K. G., & Lawley, D. N. New methods in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology*, 1968, 21, 85-96.
- Jöreskog, K. G., & Sörbom, D. *EFAP: Exploratory Factor Analysis Program*. Chicago: National Educational Resources, Inc., 1976.
- Jöreskog, K. G., & Sörbom, D. *LISREL IV—A general computer program for estimation of linear structural equation systems by maximum likelihood methods, User's Guide*. Uppsala, Sweden: University of Uppsala, 1978.
- Mosteller, F., & Tukey, J. W. *Data analysis and regression*. Reading, Mass.: Addison-Wesley, 1977.
- Mulaik, S. A. *Foundations of factor analysis*. San Francisco, Cal.: McGraw-Hill, 1972.
- Rummel, R. J. *Applied factor analysis*. Evanston, Ill.: Northwestern University Press, 1970.
- Tucker, L. R., & Lewis, C. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 1973, 38, 1-10.
- Van Driel, O. P. On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika*, 1978, 43, 225-244.

## Authors

- KROONENBERG, PIETER M. Address: Vakgroep Wijsgerige en Empirische Pedagogiek, Rijksuniversiteit Leiden, Schuttersveld 9, 2316 GX Leyden, The Netherlands. Title: Lecturer and Research Associate. Degrees: Kandidaats (B.S.) in Mathematics and Astronomy, Doctorandus (M.S.) in Mathematics, University of Leyden. Specialization: Applied Statistics.
- LEWIS, CHARLES. Address: Vakgroep Statistiek en Meettheorie, FSW, Oude Boteringestr. 23, 9712 GC Groningen, The Netherlands. Title: Lecturer and Research Associate. Degrees: B. A. in Mathematics, Swarthmore College; Ph.D. in Statistics, Princeton University. Specialization: Bayesian Statistics, Analysis of Variance, Psychometrics.