TEACHER'S CORNER

Hierarchical Regression Analysis in Structural Equation Modeling

Peter F. de Jong

Department of Psychology and Pedagogics
Vrije Universiteit

In a hierarchical or fixed-order regression analysis, the independent variables are entered into the regression equation in a prespecified order. Such an analysis is often performed when the extra amount of variance accounted for in a dependent variable by a specific independent variable is the main focus of interest (e.g., Cohen & Cohen, 1983). For example, in the area of reading achievement, there is a general interest in the specific abilities that predict reading development. Because these specific abilities are often correlated with more general abilities, such as verbal intelligence, the latter abilities are controlled for first (e.g., Wagner, Torgesen, & Rashotte, 1994). An additional reason for performing a hierarchical regression analysis is that, in these research applications, as well as in many others, the independent variables are often highly correlated. When correlated independent variables are included simultaneously in the regression model, multicollinearity arises (Cohen & Cohen, 1983). Though regularly used with observed variables, hierarchical regression analysis has not been performed with latent variables. In most applications of structural equation modeling (SEM), the latent predictors have been entered simultaneously into the regression model, although in several cases hierarchical regression analysis would have been the more appropriate approach (e.g., Guthrie et al., 1998; Normandeau & Guay, 1998; Wagner et al., 1994; Wagner et al., 1997).

In this article we describe how a hierarchical regression analysis may be conducted in SEM. The main procedure proposed is to perform a Cholesky or triangular decomposition of the intercorrelations among the latent predictors (Harman, 1976; Loehlin, 1996). First the procedure is described and then an example of a hierarchical regression analysis with latent variables is given.

THE CHOLESKY APPROACH

In Figure 1a, a path diagram of a structural model is shown in which three latent predictors, F_1 to F_3, affect a fourth latent variable F_4. The predictors F_1 to F_3 are correlated. Figure 1b depicts a Cholesky decomposition of the intercorrelations among the predictors using the Bentler–Weeks notation, which is applied in the program EQS (Bentler, 1993). The three latent predictors load in a specific way on the three new uncorrelated latent variables, F_5 to F_7 (see Harman, 1976; Loehlin, 1996). As can be seen, F_1 loads only on F_5, F_2 is dependent on F_5 and F_6, and F_3 is determined by all three newly formed latent variables. Figure 1c is similar to Figure 1b, but in this figure the notation of the LISREL program (Jöreskog & Sörbom, 1993) is used.

Mathematically, a Cholesky decomposition of the matrix of intercorrelations among the latent predictors concerns the decomposition of this matrix into the product of a triangular matrix and its transpose (Neale & Cardon, 1992). Note that this decomposition does not alter the number of degrees of freedom of the structural model in Figure 1a. Conceptually, F_5 is equal to F_1, F_6 can be interpreted as the residual of F_2 after F_1 has been partialled out, and F_7 is the residual of F_3 after both F_1 and F_2 have been partialled out. Consequently, the standardized path coefficients of F_5 with F_4, F_6 with F_4 and F_7 with F_4 (see Figure 1b) can be interpreted, respectively, as the correlation between F_1 and F_4 (r_{14}), the semipartial correlation of F_2 and F_4 controlling for F_1 ($r_{24.1}$), and the semipartial correlation of F_3 with F_4 controlling for both F_1 and F_2 ($r_{34.12}$). Because the factors F_5 to F_7 are uncorrelated, the total proportion of variance described in F_4 is $r^2_{14} + r^2_{24.1} + r^2_{34.12}$, which is exactly the kind of variance partitioning that is accomplished in a hierarchical regression analysis (e.g., Cohen & Cohen, 1983).

The order in which the predictors are included in the analysis is specified by the pattern of factor loadings of F_1 to F_3 on the extra latent variables F_5 to F_7. In the example of Figure 1b, F_3 is included after F_1 and F_2 have been controlled for. For example, to enter the predictors in the reverse order, F_3 should be equal to F_5, F_2 should be dependent on F_5 and F_6, and F_1 should load on all newly formed latent variables.

Three additional aspects of the procedure should be mentioned. First, the procedure is not only suited for a hierarchical analysis of latent variables, but it can also be adopted when either the predictors or the dependent variable or both are observed. In addition, the procedure can also be applied when two or more dependent
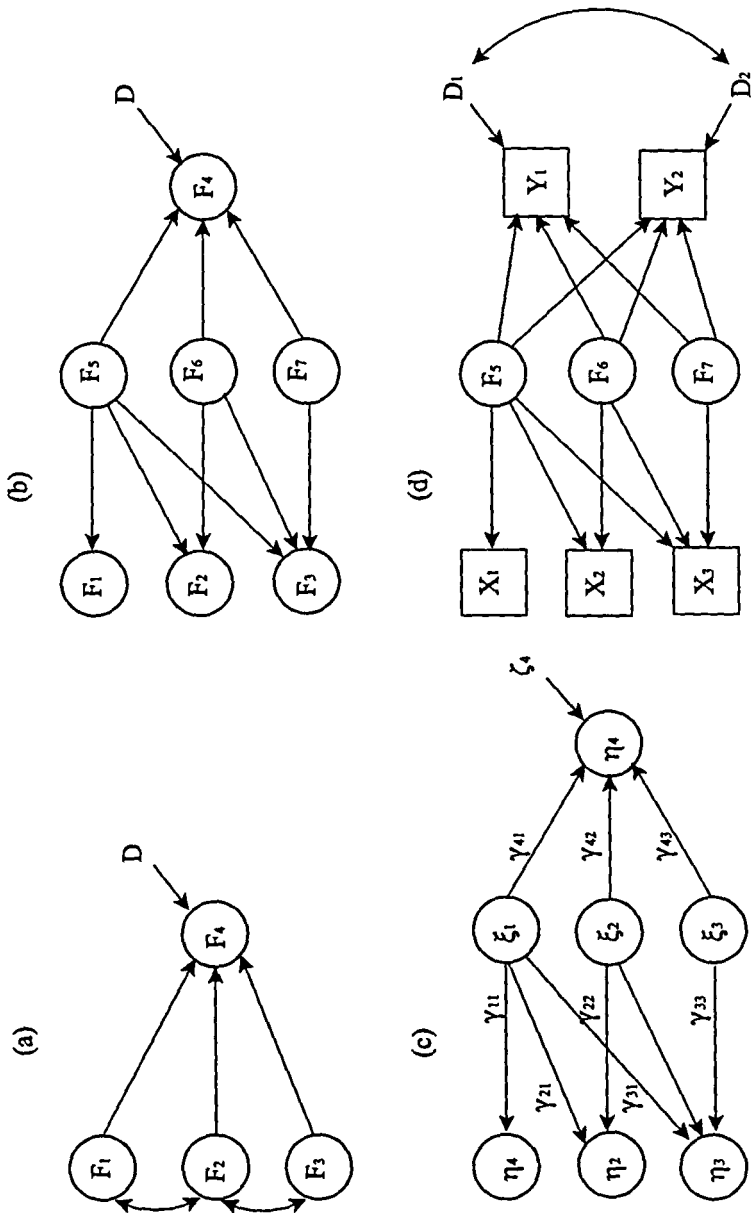
FIGURE 1 (a) Simultaneous regression model. (b) Hierarchical regression model in EQS notation. (c) Hierarchical regression model in LISREL notation. (d) Hierarchical regression model with two dependent variables. (Circles indicate latent variables; squares denote observed variables.)

variables are regressed simultaneously on the newly formed latent variables. These two aspects are illustrated in Figure 1d. Finally, it should be mentioned that the disturbance term associated with the dependent variable might be considered as another latent variable (F_8 in Figure 1b and ξ_4 in Figure 1c). This alternative specification illuminates that in the case of one dependent variable the number of original latent variables matches the number of newly formed latent variables. As a result, the procedure might be interpreted as a Cholesky decomposition of the full matrix of intercorrelations among predictors and dependent variable.[1] That is, the dependent variable F_4 loads on all newly formed latent variables (F_5–F_8), F_3 is dependent on F_5 to F_7, F_2 on F_5 and F_6, and F_2 on F_5. However, note that such an interpretation is no longer valid when two or more dependent variables are predicted simultaneously (see Figure 1d).

Several programs are available to perform hierarchical regression analysis in SEM. Currently, the most widely used are EQS (Bentler, 1993) and LISREL (Jöreskog & Sörbom, 1993). The implementation of the analysis in EQS is straightforward and can be directly derived from Figure 1b. To perform such a hierarchical regression analysis in LISREL we specify a second order factor model, Submodel 3A in the LISREL terminology (see Jöreskog & Sörbom, 1993), with the p (latent) predictors and the (latent) dependent variable as etas, which load on p extra latent variables denoted as xis (see Figure 1c where $p = 3$). The most important matrix is the matrix of factor loadings (or standardized regression coefficients) of the $p + 1$ eta variables on the p xi variables. This is a $(p + 1) \times p$ matrix, in LISREL denoted as gamma. The specific pattern of loadings in this matrix determines the order in which the (latent) predictors are entered in the regression equation. In addition, it is necessary to specify the error term of the regression of the dependent variable on the xi variables. To this extent, the last diagonal element in the psi matrix, which is a $(p + 1) \times (p + 1)$ diagonal matrix, is set free whereas the other diagonal elements are fixed to zero. Further, phi and lambda–X are identity matrices and beta and theta–delta are zero matrices. Finally, lambda–Y specifies the loadings of the observed variables, not included in Figure 1a to 1c, on the $p + 1$ eta variables (see the Appendix for an example of a setup in LISREL).[2]

ILLUSTRATIVE APPLICATION

To demonstrate the use of hierarchical regression analysis with latent variables, we analyzed part of the data reported by Wagner, Torgesen, Laughon, Simmons, and

---

[1]This alternative was suggested by an anonymous reviewer.

[2]Specification in LISREL is even somewhat easier if the disturbance term associated with the dependent variable η_4 is considered as another latent variable (ξ_4). Then, psi becomes a zero matrix and gamma is a $p \times p$ matrix (where $p = 4$). The element (p, p) of the gamma matrix is the square root of the proportion of unexplained variance in the dependent variable.

Rashotte (1993). In their kindergarten sample Wagner et al. (1993) did a confirmatory factor analysis that included (a) several phonological factors, (b) a factor that reflected more general cognitive abilities, and (c) a word recognition factor. From their factor intercorrelation matrix (see Wagner et al. [1993], Table 7, p. 97) it is evident that the phonological factors are highly correlated and that all factors are substantially related to early reading development, that is word recognition. Wagner et al. (1993) remarked that "the level of multicollinearity among predictors prevented an analysis of the relative importance of the phonological factors in accounting for word recognition" (p. 95).

For this example, we selected three tests to indicate the factor Awareness (denoted by Wagner et al. [1993] as *Analysis*), which refers to the awareness of the constituent sounds in a word, three tests to reflect Synthesis, the ability to blend separate sounds into a word, and two tests to indicate Serial Rapid Naming, the ability to retrieve the sounds of letters or digits from long-term memory. In addition, three tests were selected to reflect nonverbal intelligence. The dependent factor, Word Recognition, was formed by one test and thus not a latent but an observed variable. The correlations among the tests and their standard deviations are presented in Table 1.

To start, we conducted a confirmatory factor analysis with three phonological factors (Awareness, Synthesis, and Serial Rapid Naming), one factor for nonverbal intelligence and one word decoding factor. Note that the factor for word decoding was equivalent to the observed variable word decoding. The model appeared to have a good fit, $\chi^2(45, N = 95) = 56.13$, $p = .12$ (Confirmatory Fit Index = .98). The intercorrelations among the factors are shown in Table 2. As in the analysis reported by Wagner et al. (1993), the correlations among the factors are high, particularly the correlation between Analysis and Synthesis.

Next, we specified a structural model in which word identification was regressed simultaneously on the phonological factors and the Nonverbal Intelligence factor. The fit of the model was not altered by this alternative specification of the intercorrelations among the latent variables. The standardized regression estimates are presented in the second column of Table 3. The only significant relation was between Serial Rapid Naming and Word Recognition. The pattern of parameter estimates clearly shows the effects of multicollinearity among the latent predictors. For example, although Awareness and Synthesis have very similar correlations with the factor Word Recognition (see Table 2), the regression weight of Awareness is about .30, but nonsignificant, and the estimate for Synthesis is −.05. It is obvious that the regression coefficient of Synthesis changes when Awareness is not included in the analysis. In addition, the simultaneous solution does not indicate the extent to which the contributions of the various phonological factors to word recognition are interchangeable and whether these contributions can be accounted for by a more general cognitive ability, such as nonverbal intelligence.

TABLE 1

TABLE 1
Correlations Among the Variables

Variable	Number[a]	1	2	3	4	5	6	7	8	9	10	11	12
Awareness													
1. Segmentation	1	.—											
2. Sound isolation	3	.56	.—										
3. Sound categorization	4	.48	.45	.—									
Segmentation													
4. Blending rime	5	.60	.53	.63	.—								
5. Blending words	6	.59	.54	.56	.81	.—							
6. Blending nonwords	7	.51	.56	.57	.76	.84	.—						
Serial rapid naming													
7. Naming digits	13	.32	.51	.29	.48	.55	.59	.—					
8. Naming letters	15	.41	.53	.36	.58	.58	.62	.80	.—				
Nonverbal intelligence													
9. Corsi blocks	18	.24	.35	.35	.34	.40	.43	.39	.34	.—			
10. Visual search	19	.46	.51	.42	.47	.50	.47	.47	.45	.53	.—		
11. Spatial relations	20	.20	.38	.21	.22	.26	.37	.25	.24	.48	.51	.—	
Reading													
12. Word recognition	21	.38	.43	.34	.44	.49	.48	.55	.51	.25	.41	.26	.—
SD		1.8	4.3	3.8	5.5	4.0	3.1	0.4	0.4	6.8	4.7	6.8	4.1

[a]Number of the variable in Wagner et al. (1993).

TABLE 2
Factor Intercorrelations

Factor	1	2	3	4
1. Awareness	.—			
2. Synthesis	.880	.—		
3. Serial naming	.647	.705	.—	
4. Nonverbal intelligence	.734	.620	.585	.—
5. Word recognition	.545	.525	.591	.456

TABLE 3
Regression Analysis Predicting Word Recognition From Awareness (AW), Synthesis
(SYN), Serial Rapid Naming (SRN), and Nonverbal Intelligence (NI)

		Hierarchical Analysis					
SIP	β	Order 1	β	ΔR^2	Order 2	β	ΔR^2
AW	.30	NI	.46*	.21	NI	.46*	.21
SYN	−.05	AN	.32*	.10	SRN	.41*	.17
SRN	.42*	SYN	.10	.01	SYN	.11	.01
NI	.02	SRN	.29*	.08	AN	.12	.01
Total R^2	.40			.40			.40

Note.   SIP = Simultaneous inclusion of predictors in the regression model; ΔR^2 = incremental proportion of variance described in Word Recognition.
*$p < .01$.

To answer these latter questions, two hierarchical regression analyses were conducted in which the latent factors were entered in a prespecified order. In the first analysis the order was (a) Nonverbal Intelligence, (b) Awareness, (c) Synthesis, and (d) Serial Rapid Naming. The LISREL and EQS input files for this analysis are displayed in the Appendix. In the second analysis, the order of the phonological factors was reversed. The results of both analyses are reported in Table 3. For the hierarchical analysis the standardized regression parameters, the betas in this table, reflect a correlation with word recognition for the first factor (Nonverbal Intelligence), and semipartial correlations for the factors that are included subsequently. The square of the beta parameter reflects the incremental proportion of variance (ΔR^2) described in Word Recognition after inclusion of the factor. Thus, Nonverbal Intelligence describes 21% of the variance in word recognition. Subsequent inclusion of Awareness (Order 1) adds 10% of variance. Synthesis had no significant additional effect on word recognition after Nonverbal Intelligence and Awareness had been incorporated, whereas Serial Rapid Naming accounted for a significant 8% of the variance. Reversing the order of inclusion of the phonological factors (Order 2) revealed that neither Synthesis nor Awareness

had an effect on word recognition after controlling for Nonverbal Intelligence and Serial Rapid Naming.

This example has shown that, as known, the regression coefficients in a multiple regression analysis can be difficult to interpret when there is strong multicollinearity among the predictors. In such a situation hierarchical regression analysis should be favored. The example demonstrates how a hierarchical regression analysis can be performed within SEM and how to interpret the results.

DISCUSSION

Hierarchical regression analysis can be implemented in SEM by a Cholesky factoring of the predictors. Such a factoring does not alter model fit and does not affect the measurement part of the model.

The previous example concerned an observed variable that was regressed on several latent predictors. However, the Cholesky approach can be applied to both latent and observed variables. In addition, it is possible to use more than one dependent variable and to conduct a hierarchical regression analysis in several groups simultaneously. In the former case one can test whether the specific effect of a predictor is similar on both dependent variables (for an example of such a problem, see Gottardo, Stanovich, & Siegel, 1996). Hierarchical regression analysis in multiple groups offers, when means are taken into account, a test of group differences on a specific variable after other variables have been controlled for, which comes down to conducting an analysis of covariance.

As stated, one reason to apply hierarchical regression analysis is to examine the contributions of specific variables after controlling for more general variables. Gustafsson and Balke (1993) have described two other models that could serve the same purpose. One model is a hierarchical factors (HF) model, in which all latent predictors load on a second order factor. In this model the variance of each latent predictor is partitioned into variance due to the second order factor and residual variance. The latter can be conceived as the variance of a specific factor. An example of an HF model for three latent predictors, F_1 to F_3, is displayed in Figure 2a. As in Figure 1, we also included a dependent variable F_4 in this figure. The three latent predictors load on four uncorrelated additional latent variables. Of these variables F_5 is a second order factor that affects all three latent predictors. F_6 to F_8 are residuals associated to F_1 to F_3, respectively. However, these residuals are represented here as latent factors. An HF model for the independent variables provides direct information about the relative contributions of the general second order factor (F_5) and the independent specific factors (F_6–F_8) on one (or more) dependent variables.

In fact, Gustafsson and Balke (1993) mentioned a special type of HF model in which one of the first order factors was perfectly correlated with the second order
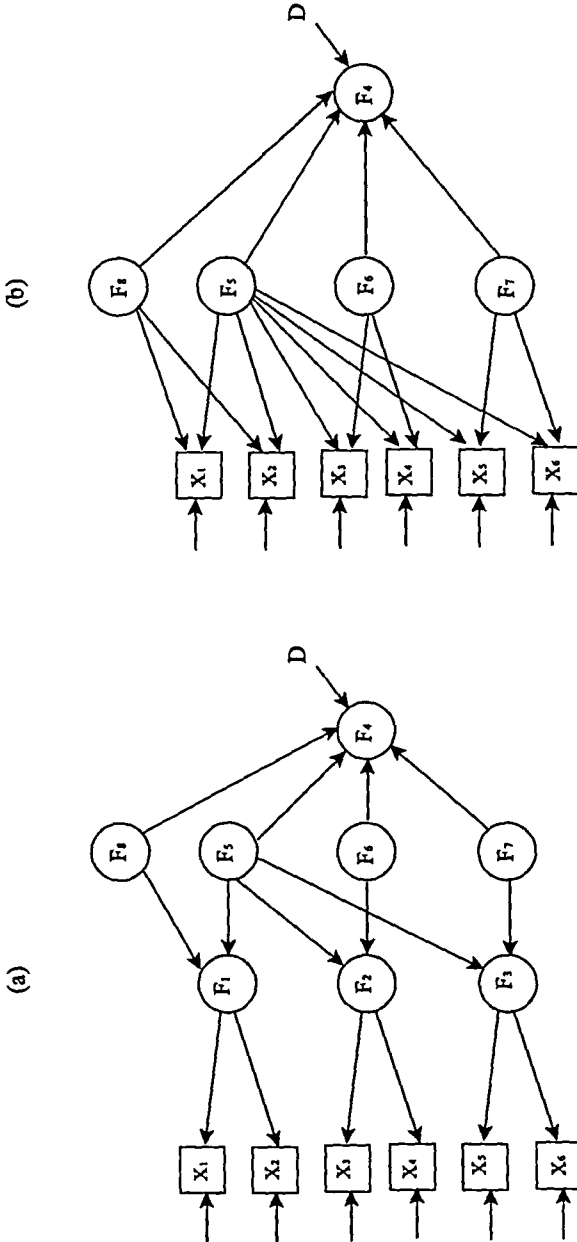
FIGURE 2   (a) Hierarchical factors model in EQS notation. (b) Nested factors model in EQS notation. (Circles indicate latent variables; squares denote observed variables.)

factor (see also Mulaik & Quartetti, 1997). The model shown in Figure 2a can be turned into this specific HF model by omitting one of the specific factors. If, for example, F_8 is omitted, F_1 is set equal to F_5. In addition, in this model the relation between the other latent predictors, F_2 and F_3, is completely due to their relation with F_1, being equal to the second order factor F_5. Note now that this specific HF model can be easily turned into the model shown in Figure 1b by including the regression of F_3 to F_6. This path represents the residual covariance between F_2 and F_3 after F_5 has been taken into account. Thus, a Cholesky decomposition of latent predictors is a less constrained version of the specific HF model proposed by Gustafsson and Balke.

Because Gustafsson and Balke (1993) considered the HF model to be fairly complicated, they proposed the nested factors (NF) model as an alternative. This model consists of a general factor and several specific factors, which are assumed to be uncorrelated with the general factor and among themselves. Regression of a dependent variable on these factors reveals the independent contributions of the general and the specific factors. The general factor is meant to replace the second order factor in the HF model. However, the general factor in the NF model is assumed to affect each observed predictor, whereas the second order factor in the HF model influences all latent predictors. Accordingly, in the NF model, the variance of each observed predictor, instead of each latent predictor, is partitioned into a part due to the general factor and a part accounted for by a specific factor. An example of a model with three latent predictors and a dependent variable is given in Figure 2b.

Gustafsson and Balke (1993) introduced the NF model as an alternative for the rather complicated HF model. However, the NF model has several disadvantages. One is that the postulate of uncorrelated specific factors is a very strong assumption. In addition, as recently noted by Mulaik and Quartetti (1997), it appears to be difficult to test this assumption in the model. Another disadvantage of the NF model is that the interpretation of the general factor is dependent on the battery of tests included in a particular study. In addition, the interpretation of this general factor might be rather difficult. The latter disadvantage also applies to the HF model, because it might prove difficult to tell what the second order factor—that is the common variance of several first order factors—reflects.

Cholesky factoring represents a less constrained version of a specific type of HF model (see previous). However, the interpretation of at least the most general Cholesky factor is straightforward. Therefore, Cholesky factoring might be considered as a better alternative to a HF model than the NF model as a means of separating the independent contributions of general and specific factors in a regression analysis.

ACKNOWLEDGMENTS

REFERENCES

Bentler, P. M. (1993). *EQS structural equations program manual.* Los Angeles: BMDP Statistical Software.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gottardo, A., Stanovich, K. E., & Siegel, L. S. (1996). The relationship between phonological sensitivity, syntactic processing, and verbal working memory in the reading performance of third-grade children. *Journal of Experimental Child Psychology, 63,* 563–582.

Gustafsson, J. E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28,* 407–434.

Guthrie, J. T., Vanmeter, P., Hancock, G. R., Alao, S., Anderson, E., & Mccan, A. (1998). Does concept-oriented reading instruction increase strategy use and conceptual learning from text? *Journal of Educational Psychology, 90,* 261–278.

Harman, H. H. (1976). *Modern factor analysis* (3rd ed., rev.). Chicago: University of Chicago Press.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: A guide to the program and applications.* Chicago: Scientific Software.

Loehlin, J. C. (1996). The Cholesky approach: A cautionary note. *Behavior Genetics, 26,* 65–69.

Mulaik, S. A., & Quartetti, D. A. (1997). First order or higher order general factor? *Structural Equation Modeling, 4,* 193–211.

Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families.* Dordrecht, The Netherlands: Kluwer.

Normandeau, S., & Guay, F. (1998). Preschool behavior and first-grade school achievement: The mediational role of cognitive self-control. *Journal of Educational Psychology, 90,* 111–121.

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85,* 83–103.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology, 30,* 73–87.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., Donahue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33,* 468–479.

EQS INPUT FILE
/TITLE
HIERARCHICAL REGRESSION ANALYSIS OF WAGNER ET AL. (1993) DATA: ORDER 1
/SPECIFICATIONS
VARIABLES = 12; CASES = 95;
METHODS = ML;
MATRIX = COR; AN = COV;
/LABELS
v1 = 1seg; v2 = 3sound; v3 = 4scateg; v4 = 5blrime; v5 = 6blword; v6 = 7blnword; v7 = 13rands; v8
  = 15ranls; v9 = 18corsi; v10 = 19search; v11 = 20sparel; v12 = 21wordr;
f1 = AW; f2 = SYN; f3 = SNAME; f4 = IQ; f5 = WD;
f6 = IQp; f7 = AWp; f8 = SYNp; f9 = SNAMEp;
/EQUATIONS
v1 = f1 + e1;
v2 = *f1 + e2;
v3 = 1.5*f1 + e3;
v4 = .5*f2 + e4;
v5 = f2 + e5;
v6 = 1.1*f2 + e6;
v7 = f3 + e7;
v8 = 1.2*f3 + e8;
v9 = 1*f4 + e9;
v10 = 1.2* f4 + e10;
v11 = f4 + e11;
!Below we specify the Cholesky decomposition
f4 = *f6;
f1 = *f6 + *f7;
f2 = *f6 + *f7 + *f8;
f3 = *f6 + *f7 + *f8 + *f9;
!This is the regression of v12 on the Cholesky factors
v12 = *f6 + *f7 + *f8 + *f9 + d12;
/VARIANCES
d12 = 8*;
f6 to f9 = 1;
e1 = 8*; e2 = 4*; e3 = 13*;
e4 = 1.5*; e5 = 1*; e6 = 3*;
e7 = .07*; e8 = .01*;
e9 = 53*; e10 = 19*; e11 = 20*;
/COVARIANCES
/MATRIX
!Insert here correlation matrix presented in Table 1
/STANDARD DEVIATION

(Continued)

!Insert here standard deviations presented in Table 1
/LMTEST
/END
LISREL 8 INPUT FILE
HIERARCHICAL REGRESSION ANALYSIS OF WAGNER ET AL. (1993) DATA: ORDER 1
da ni = 12 no = 95 ma = cm
la
1seg 3sound 4scateg 5blrime 6blword 7blnword 13rands 15ranls 18corsi 19search 20sparel 21wordr
km
!Insert here correlation matrix presented in Table 1
SD
!Insert here standard deviations presented in Table 1
mo ny = 12 ne = 5 nk = 4 ly = fu,fi ga = fu,fi ph = di,fi te = di,fi ps = di,fi
ma ly
1 0 0 0 0
2.4 0 0 0 0
2 0 0 0 0
0 1.5 0 0 0
0 1 0 0 0
0 .8 0 0 0
0 0 1 0 0
0 0 1.2 0 0
0 0 0 1 0
0 0 0 1.2 0
0 0 0 1 0
0 0 0 0 1
ma te
1 8 7 7 2 1 .07 .01 26 6 9 0
!In the first four rows of GA we specify the Cholesky decomposition
!The last row of GA is the regression of Word Recognition on the Cholesky factors
ɣa ga
1 1 0 0
1 1 1 0
1 1 1 1
1 0 0 0
1 1 1 1
fr ly 2 1 ly 3 1 ly 4 2 ly 6 2 ly 8 3 ly 9 4 ly 10 4
st 1 ly 1 1 ly 5 2 ly 7 3 ly 11 4 ly 12 5
st 1.29 ga 1 2
st 3.71 ga 2 3
st .36 ga 3 4
st 4 ga 4 1
st 1 ps 5 5
fr te 1 1 te 2 2 te 3 3 te 4 4 te 5 5 te 6 6 te 7 7 te 8 8

(Continued)

210

```
fr te 9 9 te 10 10 te 11 11
st 1 ph 1 1 ph 2 2 ph 3 3 ph 4 4
fr ps 5 5
ou ml sc ns ad = off
```

Note. The numbers in the acronyms of the variable labels correspond to the numbers of the tests in the article by Wagner et al. (1993). The start values might not be appropriate to fit alternative models to these data.