MULTILEVEL MULTIVARIATE AND STRUCTURAL EQUATION MODELS: SOME MISSING LINKS

Joop J. Hox

Utrecht University, the Netherlands

At present, multilevel structural equation models use mostly a technique developed by Muthen (1989, 1994). This technique, which is by now available in the programs Mplus, EQS, and in the preprocessor STREAMS works by decomposing the sample data into a pooled within groups and a scaled between groups covariance matrix, and analyzing these simultaneously using multi-group modeling. Since the between groups covariance matrix estimates a composite of the population within and between matrices, the analysis model tends to become complicated in this approach. Also, to keep the analysis manageable, the analysis neglects the fact that with unbalanced groups, the between groups covariance matrix reflects a mixture of populations. Various simulations have shown that with reasonable sample sizes at both levels, ignoring the unbalance does not produce sizeable bias. A different approach, suggested by Goldstein is to use a conventional multilevel regression program as a preprocessor to produce a direct maximum likelihood estimate of the population within and between covariance matrices. This paper examines the advantages and disadvantages of the two approaches, using example data.

Key words: SEM, hierarchical model, two-level SEM

A general analysis approach that includes both factor and path analysis is *Structural Equation Modeling*, or *SEM*. The interest in SEM is often on *theoretical constructs*, which are represented by the latent factors. The factor model, which is often called the measurement model, specifies how the latent factors are measured by the observed variables. The relationships between the theoretical constructs are represented by regression or path coefficients between the factors. The structural equation model implies a structure for the covariances between the observed variables, which explains the alternative name *Structural equation Analysis*. However, the model can be extended to include means of observed variables or factors in the model, which makes structural equation modeling a more accurate name.

Structural equation models for multilevel data have been elaborated, among others, by Goldstein and McDonald (Goldstein & McDonald, 1988; McDonald & Goldstein, 1989, McDonald, 1994), Muthén and Satorra (Muthén, 1989; Muthén & Satorra, 1989) and Longford and Muthén (Longford & Muthén, 1992). I refer to McArdle and Hamagami (1996) for a comparison between multilevel regression techniques and standard multigroup SEM. The

approach by Muthén is particularly interesting, because he shows that structural equation modeling (SEM) of multilevel data is possible using available SEM software, such as Lisrel (Jöreskog & Sörbom, 1996), Eqs, or Amos (Arbucle & Wothke, 1999). For an introductory exposition of Muthén's method, see Muthén (1994), Hox (1995), Kaplan and Elliot (1997) and Li, Duncan, Harmer, Acock, and Stoolmiller (1998). Meanwhile, software has appeared that includes these multilevel extensions directly in the SEM program (M*plus*, see Muthén & Muthén, 1998; Eqs 6.0, as promised in spring 2000) or acts as a front end for conventional SEM software (Streams, see Gustaffson & Stahl, 1999). Heck and Thomas (2000) present an extended example of multilevel SEM, which uses Muthén's method and discusses the implementation details for the programs Lisrel, Streams, and Mplus.

This paper discusses two different approaches to multilevel SEM: the approach originally proposed by Muthén, and an approach that is based on direct estimation of the covariance matrices at the distinct levels, as proposed by Goldstein (1987, 1995) and applied by Rowe and Hill (1998).

1. DECOMPOSING MULTILEVEL VARIABLES

Multilevel structural models assume that we have a population of individuals that are divided into groups. The individual data are collected in a *p*-variate vector $\mathbf{Y}_{ig}$ (subscript *i* for individuals, *g* for groups). Cronbach and Webb (1975) propose to decompose the individual data $\mathbf{Y}_{ig}$ into a between groups component $\mathbf{Y}_B = \overline{\mathbf{Y}}_g$, and a within groups component $\mathbf{Y}_B = \mathbf{Y}_{ig} - \overline{\mathbf{Y}}_g$. In other words, for each individual we replace the observed *Total* score $\mathbf{Y}_T = \mathbf{Y}_{ig}$ by its components: the group component $\mathbf{Y}_B$ (the disaggregated group mean) and the individual component $\mathbf{Y}_W$ (the individual deviation from the group mean.) These two components have the attractive property that they are orthogonal and additive (cf. Searle, Casella & McCulloch, 1992):

$$\mathbf{Y}_T = \mathbf{Y}_B + \mathbf{Y}_W \qquad (1.1)$$

This decomposition can be used to compute a between groups covariance matrix $\mathbf{\Sigma}_B$ (the population covariance matrix of the disaggregated group means $\mathbf{Y}_B$) and a within groups covariance matrix $\mathbf{\Sigma}_W$ (the population covariance matrix of the individual deviations from the group means $\mathbf{Y}_W$). These covariance matrices are also orthogonal and additive:

$$\mathbf{S}_T = \mathbf{S}_B + \mathbf{S}_W \qquad (1.2)$$

Following the same logic, we can also decompose the sample data. Suppose we have data from *N* individuals, divided into *G* groups (subscript *i* for individuals, *i=1…N*; subscript *g* for groups, *g=1…G*). If we decompose the sample data, we have for the sample covariance matrices:

$$\mathbf{S}_T = \mathbf{S}_B + \mathbf{S}_W \tag{1.3}$$

Multilevel structural equation modeling assumes that the population covariance matrices $\mathbf{\Sigma}_B$ and $\mathbf{\Sigma}_W$ can be described by separate models for the between groups and within groups structure. To estimate the model parameters, the factor loadings, path coefficients, and residual variances, we need maximum likelihood estimates of the population between groups covariance matrix $\mathbf{\Sigma}_B$ and the population within groups covariance matrix $\mathbf{\Sigma}_W$. What we have is the sample between groups matrix $\mathbf{S}_B$ and the sample within groups matrix $\mathbf{S}_W$. Unfortunately, we cannot simply use $\mathbf{S}_B$ as an estimate of $\mathbf{\Sigma}_B$, and $\mathbf{S}_W$ for $\mathbf{\Sigma}_W$. The situation is a bit more complicated.

1.2 MUTHÉN'S PSEUDOBALANCED APPROACH

In the special case of balanced groups, estimation is straightforward (Muthén, 1989). If we have *G* balanced groups, with *G* equal group sizes *n* and a total sample size *N=nG*, we define two sample covariance matrices: the pooled within covariance matrix $\mathbf{S}_{PW}$ and the scaled between covariance matrix $\mathbf{S}^*_B$.

Muthén (1989) shows that an unbiased estimate of the population within groups covariance matrix $\mathbf{\Sigma}_W$ is given by the pooled within groups covariance matrix $\mathbf{S}_{PW}$, calculated in the sample by:

$$\mathbf{S}_{PW} = \frac{\sum_{g}^{G}\sum_{i}^{n}(Y_{gi} - \bar{Y}_g)(Y_{gi} - \bar{Y}_g)'}{N - G} \tag{1.4}$$

Equation (1.4) corresponds to the conventional equation for the covariance matrix of the individual deviation scores, with *N-G* in the denominator instead of the usual N-1.

Since the pooled within groups covariance matrix $\mathbf{S}_{PW}$ is an unbiased estimate of the population within groups covariance matrix $\mathbf{\Sigma}_W$, we can estimate the population within group structure by constructing and testing a model for $\mathbf{S}_{PW}$.

The scaled between groups covariance matrix for the disaggregated group means $\mathbf{S}_{B,}$ calculated in the sample, is given by:

$$\mathbf{S}_B^* = \frac{\sum_g^G n(\bar{Y} - \bar{Y}_g)(\bar{Y} - \bar{Y}_g)'}{G-1} \qquad (1.5)$$

Muthén (1989, 1990) shows that $\mathbf{S}_{PW}$ is the maximum likelihood estimator of $\mathbf{\Sigma}_W$, with sample size *N-G*, and $\mathbf{S}^*_B$ is the maximum likelihood estimator of the composite $\mathbf{\Sigma}_W + c\mathbf{\Sigma}_B$, with sample size *G*, and *c* equal to the common group size *n*:

$$\mathbf{S}_{PW} = \hat{\Sigma}_W \qquad (1.6)$$

and

$$\mathbf{S}_B^* = \hat{\Sigma}_W + c\hat{\Sigma}_B \qquad (1.7)$$

Equations 1.6 and 1.7 suggest using the multi-group option of conventional SEM software for a simultaneous analysis at both levels. However, if we model the between groups structure, we cannot simply construct and test a model for $\mathbf{S}_B$, because $\mathbf{S}_B$ estimates a combination of $\mathbf{\Sigma}_W$ and $\mathbf{\Sigma}_B$. Instead, we have to specify for $\mathbf{S}_B$ two models: one for the within groups structure and one for the between groups structure. The procedure is that we specify two groups, with covariance matrices $\mathbf{S}_{PW}$ and $\mathbf{S}_B$ (based on N-G and G observations). The model for $\mathbf{\Sigma}_W$ must be specified for both $\mathbf{S}_{PW}$ and $\mathbf{S}_B$, with equality restrictions between both `groups' to guarantee that we are indeed estimating the same model in both covariance matrices, and the model for $\mathbf{\Sigma}_B$ is specified for $\mathbf{S}_B$, with the scale factor *c* built into the model.

The reasoning given above applies only in the so-called *balanced* case, that is, if all groups have the same group size. In the balanced case, the scale factor *c* is equal to the common group size *n*. The unbalanced case, where the group sizes differ, with *G* groups of unequal sizes, is more complicated. In this case, $\mathbf{S}_{PW}$ is still the maximum likelihood estimator of $\mathbf{\Sigma}_W$, but $\mathbf{S}^*_B$ now estimates a different expression for each set of groups with distinct group size *d*:

$$\mathbf{S}_{Bd}^* = \hat{\Sigma}_W + c_d\hat{\Sigma}_B \qquad (1.8)$$

where equation 1.8 holds for each distinct set of groups with a common group size equal to n_d, and c_d=n_d (Muthén, 1990, 1994). Full Information Maximum Likelihood (FIML) estimation for unbalanced groups implies specifying a separate between-group model for each distinct group

size. These between groups models have different scaling parameters c_d for each distinct group size, and require equality constraints across all other parameters and inclusion of a mean structure (Muthén, 1994, p. 385). Thus, using conventional SEM software for the unbalanced case requires a complicated modeling scheme that creates a different 'group' for each set of groups with the same group size. This results in large and complex models, with possibly groups with a sample size less than the number of elements in the corresponding covariance matrix. This makes full Maximum Likelihood estimation problematic, and therefore Muthén (1989, 1990) proposes to ignore the unbalance, and to compute a single $\mathbf{S}^*_B$. The model for $\mathbf{S}^*_B$ includes an ad hoc estimator c^* for the scaling parameter, which is close to the average sample size:

$$c^* = \frac{N^2 - \sum_g^G n_g^2}{N(G-1)} \tag{1.9}$$

This solution is not a full maximum likelihood solution. The result is a Limited Information Maximum Likelihood (LIML) solution, which McDonald (1994) calls a pseudobalanced solution. Muthén (1989, 1990) shows that $\mathbf{S}^*_B$ is a consistent and unbiased estimator of the composite $\mathbf{\Sigma}_W + c\mathbf{\Sigma}_B$. This means that with large samples (of both individuals *and* groups!) $\mathbf{S}_B$ generally becomes a close estimate of $\mathbf{\Sigma}_B$, and the pseudobalanced solution should produce a good approximation given adequate sample sizes.

Since $\mathbf{S}_B$ is not a maximum likelihood estimator, the analysis produces only approximate parameter estimates and standard errors. However, when the group sizes are not extremely different, the pseudobalanced estimates will be close enough to the full maximum likelihood estimates to be useful in their own right. Comparisons of pseudobalanced estimates with full maximum likelihood estimates or with known population values have been made by Muthén (1990, 1994), Hox (1993), and McDonald (1994). Their main conclusion is that the pseudobalanced parameter estimates are fairly accurate and useful for a variety of multilevel problems. A large simulation study by Hox and Maas (2000) assesses the robustness of the pseudobalanced method against unequal groups, small sample sizes at both the individual and the group level, in the presence of a low or a high intraclass correlation (ICC). In this study, the within groups part of the model poses no problems. The most important problem in the between groups part of the model, is the occurrence of inadmissible estimates, when the group level sample size is small (50) and the intracluster correlation is low. When an admissible solution is

found, the factor loadings are generally accurate. However, the residual variances are underestimated, and the standard errors are generally too small. Having more or larger groups or a higher ICC does not effectively compensate this. Therefore, while the nominal alpha level is 5%, the operating alpha level is about 8% in all simulated conditions with unbalanced groups. The strongest contributing factor is an inadequate sample size at the group level. Imbalance is also a problem for the overall goodness-of-fit test. For balanced data, the chi-square test for goodness-of-fit is accurate. For unbalanced data, the model is rejected too often, which results in an operating alpha level of about 8%. The size of the biases is comparable to the effect of moderate non-normality in ordinary modeling. Hox and Maas conclude that the approximate solution is useful, provided the group level sample size is at least 100, and keeping in mind that the operating alpha level is somewhat higher than the nominal alpha level.

The multilevel part of the structural equation model outlined above is simpler than that of the multilevel regression model. It is comparable to the multilevel regression model with random variation of the intercepts. There is no provision for randomly varying slopes (factor loadings and path coefficients). Although it would be possible to include cross-level interactions, introducing interaction variables of any kind in structural equation models is neither simple nor elegant (cf. Bollen, 1989). An interesting approach would be to allow for different within groups covariance matrices in different subsamples.

The pseudobalanced approach needs the pooled within and the scaled between covariance matrices. Standard software does not provide these directly. One solution is to use special software that calculates these matrices directly, such as the freeware program SPLIT2 (Hox, 1994) or the preprocessor STREAMS (Gustaffson & Stahl, 1999). A different solution is to use standard software such as SPSS to calculate the correlations and standard deviations of the deviation scores and the disaggregated means. The correlations are scale-free numbers, and therefore the correct ones. The standard deviations will be calculated using *N-1* in the denominator, instead of the denominators *N-g* for the pooled within covariances and *G-1* for the scaled between groups covariances. In other words, the standard deviations of the pooled within matrix must be corrected by multiplying them by $\sqrt{N-1/N-G}$, and the standard deviations of the scaled between matrix must be corrected by multiplying them by $\sqrt{N-1/G-1}$. It is easy to make these corrections by hand, and then input the correlations with the corrected standard deviations into a standard SEM program.

Since the pseudobalanced approach needs the within groups model both for the pooled within groups and the scaled between groups model, and needs to incorporate the scaling factor

for the between groups model, the actual model can become quite complicated. In addition, some software has difficulties finding good starting values. Several software writers have addressed these problems. The program STREAMS (Gustaffson & Stahl, 1999) acts as a preprocessor for standard SEM software. For two-level SEM, it calculates the pooled within and scaled between matrices, and writes the complicated setup, including starting values based on previous analyses. The program Mplus (Muthén & Muthén, 1998) hides all the complications of the pseudo-balanced approach from the user. It also uses by default robust estimators for the standard errors and adjusts the chi-square test statistic for the heterogeneity that results from mixing groups of different sizes (cf. Muthén & Satorra, 1995).

1.2.1 An Example of a Multilevel Factor Analysis Using the Pseudobalanced Approach

The example data are taken from Van Peet (1992). They are the scores on six intelligence measures of 187 children from 37 families. The six intelligence measures are: word list, cards, matrices, figures, animals, and occupations. The data have a multilevel structure, with children nested within families. If intelligence is strongly influenced by shared genetic and environmental influences in the families, we may expect rather strong between family effects.

To begin, the individual scores on the six measures are decomposed into disaggregated group means and individual deviations from the group means. Table 1.1 shows the means and variances of the scores, and the Intra Class Correlation (ICC),[1] with the family and individual level variances calculated using standard formulas for the variances. Note that, within rounding errors, the family level variance and the individual level variance sum to the total variance.

[1]The ICC can be estimated by analysis of variance procedures (Hays, 1994), or from the intercept-only model using a multilevel approach. Here, it is estimated from the pooled within groups and between groups variances.

Table 1.1 Means, variances and ICC for family data

Measure	Mean	Total Variance	Family Variance	Individual Variance	ICC
Word list	29.8	15.21	7.48	7.73	.38
Cards	32.7	28.47	13.65	14.82	.36
Matrices	31.7	16.38	5.24	11.14	.16
Figures	27.1	21.23	6.84	14.38	.16
Animals	28.7	22.82	8.46	14.36	.22
Occupations	28.3	21.42	9.11	12.31	.29

The results in Table 5.1 suggest that there are indeed sizeable family effects. To analyze the factor structure of the six measures on the individual and family level, we compute the pooled within family covariance matrix $\mathbf{S}_{PW}$ and the scaled between family covariance matrix $\mathbf{S}_{B}$, using equations (1.4) and (1.5). The results are in Table 1.2 and 1.3.

Table 11.2 Pooled within families covariances and correlations

	1	**2**	**3**	**4**	**5**	**6**
Word list	9.59	*.24*	*.30*	*.20*	*.09*	*.05*
Cards	3.16	18.37	*.49*	*.14*	*.14*	*.01*
Matrices	3.49	7.83	13.81	*.14*	*.11*	*.07*
Figures	2.64	2.56	2.12	17.38	*.26*	*.19*
Animals	1.15	2.45	1.74	4.66	17.81	*.45*
Occupations	0.60	0.15	1.02	3.10	7.42	15.27

Note: italic entries in upper diagonal are the correlations

Table 11.3 Scaled between families covariances and correlations

	Wlist	Crds	Matr	Figs	Anim	Occ
Word list	38.62	*.56*	*.58*	*.19*	*.60*	*.32*
Cards	29.24	70.53	*.66*	*.49*	*.41*	*.17*
Matrices	18.74	29.02	27.08	*.47*	*.55*	*.10*
Figures	6.93	24.35	14.67	35.37	*.31*	*.13*
Animals	24.45	22.51	19.05	12.15	43.70	*.35*
Occupations	13.70	9.97	3.53	5.40	15.74	47.04

Note: italic entries in upper diagonal are the correlations

The covariances in Table 1.3 appear very large, which is what they should be, because the covariance matrix in Table 1.3 is the scaled between groups matrix. This matrix equals the within groups matrix plus the between groups matrix multiplied by the average group size. The average group size for these family data, indicated by the scaling factor in equation (1.6) is 5.04. The correlations in Tables 1.2 and 1.3 suggest that the structure is much stronger at the family level than at the individual level.

Typically, in multilevel SEM, there are many more individuals than groups, and hence the number of observations for the pooled within groups covariance matrix (N-G) is much larger than the number of observations for the between groups covariance matrix (G). In this case, the number of observations on the individual level is 187-37=150, while on the family level it is 37. Thus, it makes sense to start on the individual level by constructing a model for $\mathbf{S}_{PW}$.

An exploratory factor analysis on the correlations derived from $\mathbf{S}_{PW}$ suggests two factors, with the first three measures loading on the first factor, and the last three measures on the last. A confirmatory factor analysis on $\mathbf{S}_{PW}$ confirms this model: χ^2=7.21, df=8, p=.51. A model with just one general factor is rejected: χ^2=44.87, df=9, p=.00. Figure 1.1 presents the conventional graphic representation of the individual level (within families) model.

The next step is the specification of a family model. For this, we must analyze the matrices $\mathbf{S}_{PW}$ and $\mathbf{S}_B$ simultaneously with the multigroup procedure. First, we specify the individual model for both 'groups' with equality restrictions applied across both groups for all parameters. Next, we must specify an additional family model for $\mathbf{S}_B$.

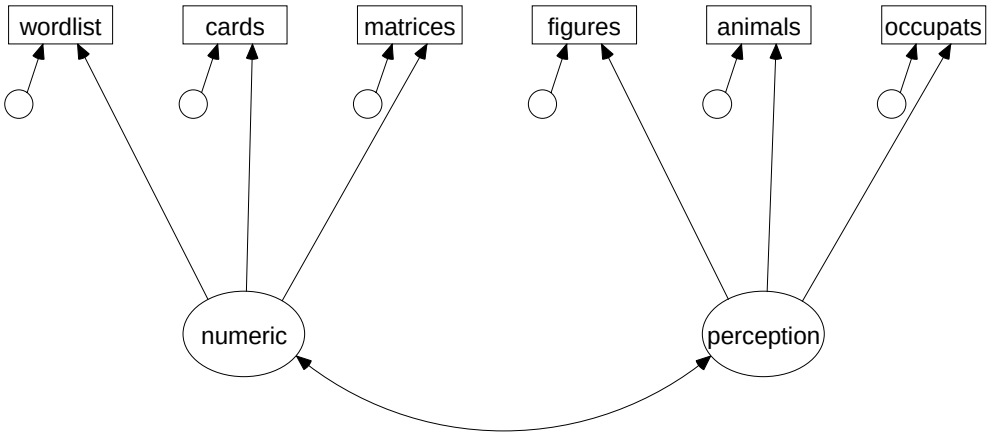

Figure 1.1 Within families model for Van Peet data

We start by estimating some 'benchmark' models, to test whether there is any between family structure at all. The simplest model is the null model that omits the specification of a family level model. If the null model holds, there is no family level structure at all; all covariances in $\mathbf{S}_B$ are the result of individual sampling variation. If this null-model holds, we may as well continue our analyses using simple single level analysis methods. The next model is the independence model, which specifies only variances on the family level, but no covariances. A graphical representation of the independence model for $\mathbf{S}_B$ is given in Figure 1.2.

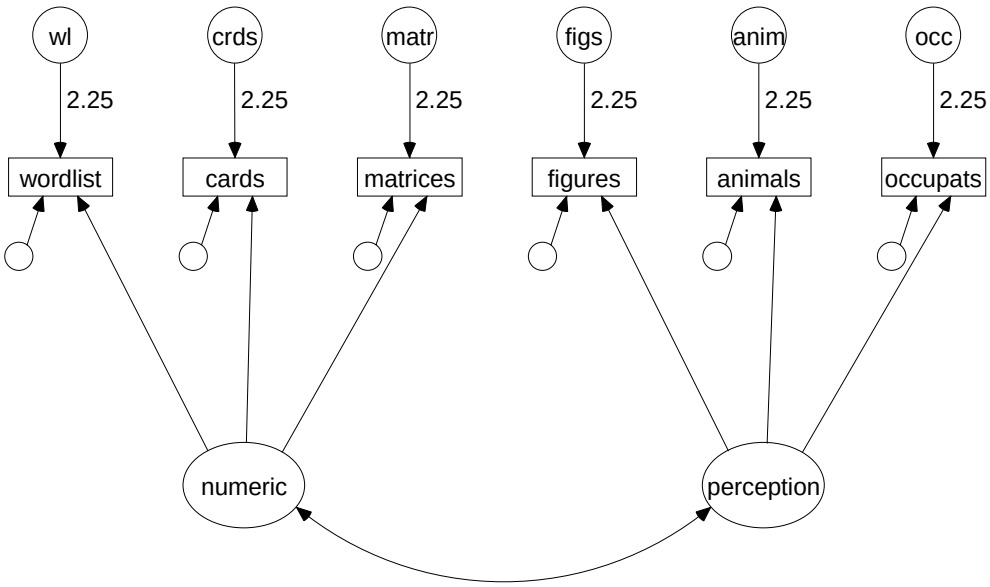

Figure 1.2 Within model + independence model for between structure, Van Peet data

Note that in Figure 1.2 I have fixed the loadings for the family level variables (the six 'factors' in the circles going from 'wl' to 'occ') not to one, as is usual, but to 2.25, which is the square root of the scale factor. This is to transform the family level variables to their proper scale. Since this is a fixed value, it has no influence on the global fit of the model, but it is necessary for a correct interpretation.

The independence model estimates for the family level structure only the variances of the family level variables 'wl' to 'occ', If the independence model holds, there is family level variance, but no substantively interesting structural model. Nevertheless, in this case it is still useful to apply multilevel analysis, because this produces unbiased estimates of the individual model parameters. Since there is no interesting between groups model in this case, we can simply analyze the pooled within matrix, at some cost in loosing information from *G* observations. If the independence model is rejected, there is some kind of structural model on the family level. To examine the best possible fit given the individual level model, we can estimate the saturated model; which fits a full covariance matrix to the family level observations. This places no restrictions on the family model.[1] Table 1.4 shows the results of estimating these models:

Table 1.5 Family level benchmark models

Family model	Chi-square	df	p
Null	125.4	29	.00
Independence	52.5	23	.00
Saturated	7.2	8	.51

Both the null model and the independence model are rejected. Next, we specify for the family level the same two models we have used for the individual level. Again, the two factor model fits well. However, on the family level a one factor model fits almost as well, as Table 5.3 shows:

[1]To establish the within groups structure, we can specify the saturated model for the between structure, and then explore the within model simultaneously in both $\mathbf{S}_{PW}$ and $\mathbf{S}_{B}$. However, since $\mathbf{S}_{PW}$ is usually based on many more observations than $\mathbf{S}_{B,}$ little information is lost by only analyzing $\mathbf{S}_{PW}$, while this makes the setups much simpler.

Table 1.6 Family level factor models

Family model	Chi-square	df	P
1 Factor	21.3	17	.21
2 Factors	20.1	16	.22

The principle of using the simplest model that fits well, leads to acceptance of the one factor model on the family level, which is depicted in Figure 1.3. This model also shows reasonable to good 'goodness-of-fit' indices: the traditional fit index GFI is 0.88, which is too low, but the comparative fit index CFI is 0.97, and the root mean square error of approximation RMSEA is 0.04, both of which are acceptable values.

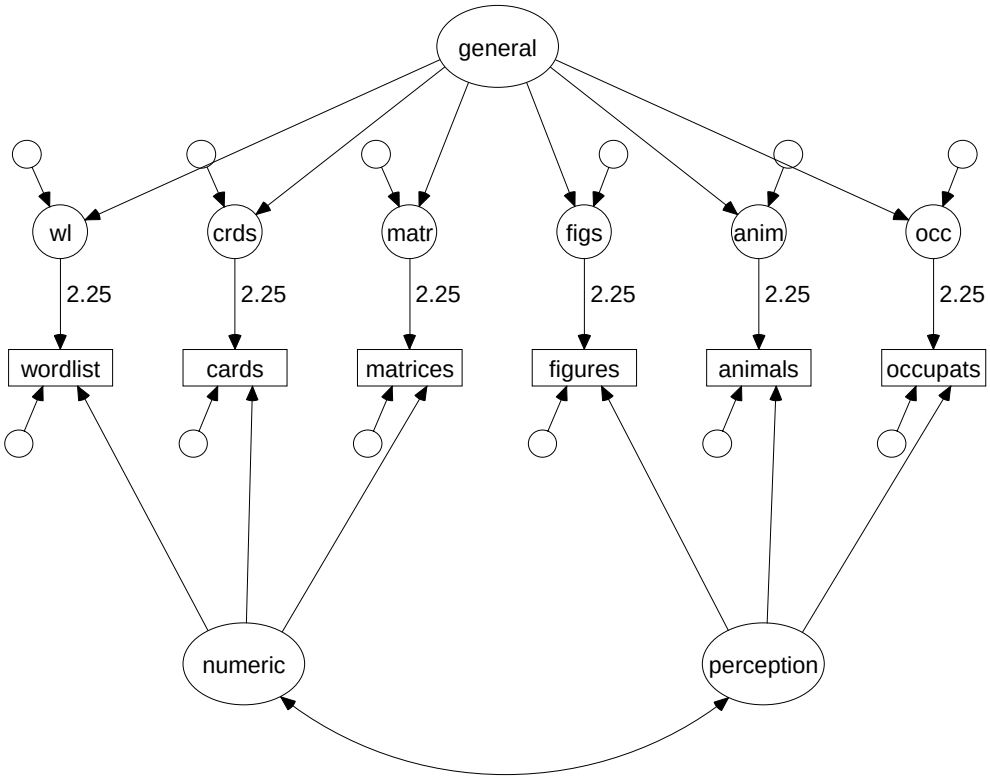

Figure 1.4 Final 2-level factor model for Van Peet data

The factor loadings and the residual error variances for the two-level factor model are in Table 1.6 below.

Table 1.6 Individual and family level estimates, standard SEM software

	Individual level			Family level	
	Numeric	**Perception**	**resid. var.**	**General**	**resid. var.**
Word list	1*		8.13 (1.03)	1*	1.79 (1.14)
Cards	2.34 (.66)		10.76 (2.44)	1.24 (.36)	3.95 (1.84)
Matrices	2.39 (.74)		5.80 (2.28)	0.82 (.22)	-.03 (.66)
Figures		.63 (.21)	15.81 (1.97)	0.56 (.27)	2.45 (1.45)
Animals		1.51 (.55)	6.70 (3.92)	0.98 (.31)	1.36 (1.33)
Occupations		1*	10.43 (2.08)	0.41 (.30)^{ns}	5.69 (1.97)

Standard errors in parentheses. Correlation between individual factors: 0.22; * = fixed; ns = not significant

There is a small anomaly in Table 1.6: one of the error variances at the family level is negative. The value is very small, and this kind of problem tends to occur when the sample sizes is small. Thirty-seven families is indeed a small sample, and the usual treatment is to fix the small negative variance estimate at zero. Table 1.6 suggests an interpretation that on the family level, where the effects of the shared genetic and environmental influences are visible, one general (*g*) factor is sufficient to explain the covariances between the intelligence measures. On the individual level, where the effects of individual idiosyncratic influences are visible, we need two factors. These results could be fitted into Cattel's (1971) theory of fluid and crystallized intelligence, which states that, as a result of individual factors (education, physical and social environment), the general g-factor `crystallizes' into specific individual competencies. Gustaffson and Stahl (1999) also analyze this data set, and arrive at a different model.

The results given above were produced using standard SEM software on the pooled within and the scaled between covariance matrix, produced by the program SPLIT2 (Hox, 1994). Using Mplus (Muthén & Muthén, 1998), we can analyze the multilevel factor model directly. If we use Mplus to produce a Maximum Likelihood solution, the results are almost identical: a chi-square of 21.7 (df=17, p=0.20) and estimates and standard errors that are very close to the estimates given in Table 1.6. However, by default Mplus produces for two-level models maximum likelihood estimates with robust standard errors and a corrected chi-square test. The corrected chi-square is 24.0 (df=17, p=0.12), which is somewhat different from the standard chi-square. The Mplus maximum likelihood estimates and robust standard errors are in Table 1.7.

Table 1.7 Individual and family level estimates, Mplus robust standard errors

	Individual level			Family level	
	Numeric	**Perception**	**resid. var.**	**General**	**resid. var.**
Word list	1*		8.13 (1.02)	1*	1.80 (1.42)
Cards	2.34 (.60)		10.75 (2.76)	1.24 (.43)	3.96 (1.55)
Matrices	2.39 (.86)		5.81 (2.51)	0.82 (.23)	-.04 (.56)
Figures		.63 (.17)	15.81 (1.51)	0.56 (.28)	2.46 (1.12)
Animals		1.51 (.55)	6.70 (3.91)	0.98 (.27)	1.37 (.88)
Occupations		1*	10.43 (1.74)	0.41 (.23)^{ns}	5.71 (2.18)

Standard errors in parentheses. Correlation between individual factors: 0.22; * = fixed; ns = not significant

The results are not greatly different from the standard asymptotic standard errors. Still, using the Mplus robust estimates, the covariance between the within-families factors, which is barely significant using the asymptotic standard error (p=0.047), is clearly significant when the corrected standard error is used (p=0.034).

1.2.2 Goodness of Fit Using the Pseudobalanced Approach

Standard SEM programs, and specialized programs like STREAMS and Mplus, produce in addition to the chi-square test a large number of goodness-of-fit indices that indicate how well the model fits the data. Statistical tests for model fit have the problem that their power varies with the sample size. If we have a very large sample, the statistical test will almost certainly be significant. Thus, with large samples, we will always reject our model, even if the model actually describes the data very well. Conversely, with a very small sample, the model will always be accepted, even if it fits rather badly.

Given the sensitivity of the chi-square statistic for sample size, researchers have proposed a variety of alternative fit indices to assess model fit. All goodness-of-fit measures are some function of the chi-square and the degrees of freedom. Most of these fit indices not only consider the fit of the model, but also its simplicity. A saturated model, that specifies all possible paths between all variables, always fits the data perfectly, but it is just as complex as the observed data. In general, there is a trade-off between the fit of a model and the simplicity of a model. Several goodness-of-fit indices assess simultaneously both the fit and the

simplicity of a model. The goal is to produce a goodness-of-fit index that does not depend on the sample size or the distribution of the data. In fact, most goodness-of-fit indices still depend on sample size and distribution, but the dependency is much smaller than that of the routine chi-square test.

Modern SEM software computes a bewildering array of goodness-of-fit indices. All of them are functions of the chi-square statistic, but some include a second function that penalizes complex models. For instance, Akaike's information criterion (AIC), is twice the chi-square statistic minus the degrees of freedom for the model. For an overview and evaluation of a large number of fit indices, including those mentioned here, I refer to Gerbing and Anderson (1993).

Jöreskog and Sörbom (1989) have introduced two goodness-of-fit indices called GFI (Goodness of Fit) and AGFI (Adjusted GFI). The GFI indicates goodness-of-fit, and the AGFI attempts to adjust the GFI for the complexity of the model. Bentler (1990) introduces a similar index called the Comparative Fit Index CFI. Two other well-known measures are the Tucker-Lewis Index TLI (Tucker & Lewis, 1973), better known as the Non-Normed Fit Index or NNFI, and the Normed Fit Index NFI (Bentler & Bonett, 1980). Both the NNFI and the NFI adjust for complexity of the model. Simulation research shows that all these indices still depend somewhat on sample size and estimation method (e.g., ML or GLS), with the CFI and the TLI/NNFI showing the best overall performance (Chou & Bentler, 1995; Kaplan, 1995). If the model fits perfectly, these fit indices should have the value 1. Usually, a value of at least 0.90 is required to accept a model, while a value of at least 0.95 is required to judge the model fit as 'good.' However, these are just rules of thumb.

A relatively modern approach to model fit is to accept that models are only approximations, and that perfect fit may be too much to ask for. Instead, the problem is to assess how well a given model approximates the true model. This view led to the development of an index called RMSEA, for Root Mean Square Error of Approximation (Browne & Cudeck, 1992). If the approximation is good, the RMSEA should be small. Typically, a RMSEA of less than 0.05 is required, and statistical tests or confidence intervals can be computed to test if the RMSEA is significantly larger than this lower bound.

Given the many possible goodness-of-fit indices, the usual advice is to assess fit by inspecting several fit indices that derive from different principles. Therefore, for the confirmatory factor model for the family data, I reported the chi-square test, and the fit indices GFI, CFI and RMSEA.

A general problem with these goodness-of-fit indices is that they are reported for the

entire model. Therefore, the goodness-of-fit is reflects both the fit in the within model and the between model. Since the sample size for the within 'group' is generally the largest, this part of the model dominates the value of the fit indices. Clearly, we would prefer to assess the fit for both parts of the model separately.

Since the within groups sample size is usually much larger than the between groups sample size, we do not loose much information if we model the within groups matrix separately, and interpret the fit indices produced in this analysis separately.

A simple way to obtain goodness-of-fit indices for the between model is to specify for the within groups matrix a saturated model. The saturated model estimates all covariances between all variables. It has no degrees of freedom, and always fits the data perfectly. As a result, the degree of fit indicated by the goodness-of-fit indices, represents the fit of the between model. This is not the best way to assess the fit of the between model, because the perfect fit of the within model also influences the value of the fit index. Fit indices that are mostly sensitive to the degree of fit will show a spuriously good fit, while fit indices that also reflect the parsimony of the model may even show a spurious lack of fit.

A better way to indicate the fit of the within and between model separately is to calculate these by hand. Most fit indices are a function of the chi-square, sample size *N*, and degrees of freedom *df*. Some consider only the model under consideration, the target model M_t, others also consider a baseline model, usually the independence model. By estimating the independence and the target model for the within matrix, with a saturated model for the between matrix, we can assess how large the contribution to the overall chi-square is for the various within models. In the same way, by estimating the independence and the target model for the between matrix, with a saturated model for the within matrix, we can assess how large the contribution to the overall chi-square is for the various between models. Using this information, we can calculate the most common goodness-of-fit indices. Most SEM software produces the needed information, and the references and formulas are in the user manuals and in the general literature (e.g. Gerbing & Anderson, 1992).

Table 1.8 gives the separate chi-squares, degrees of freedom, and sample sizes for the independence model and the final model for the family intelligence example.

Table 1.8 Chi-squares and degrees of freedom for individual and family level models separately

7	Individual level, Between model saturated		Family level, within model saturated	
	independence	**2 factors**	**independence**	**1 factor**
chi-square	148.26	7.27	45.23	13.96
Df	15	8	15	13
N	150	150	37	37

The calculation of the widely used goodness-of-fit index GFI is too complicated to be carried out by hand. The comparative fit index CFI (Bentler, 1990) is given by

$$CFI = 1 - \frac{\chi_t^2 - df_t}{\chi_I^2 - df_I} \qquad (1.10).$$

In equation (1.10), χ_t^2 is the chi-square of the target model, χ_I^2 is the chi-square for the independence model, and df_t and df_I are the degrees of freedom for the target and the independence model. If the difference of the chi-square and the degrees of freedom is negative, it is replaced by zero.

The Tucker-Lewis index, TLI, which is also known as the Non-Normed Fit Index, NNFI, is given by

$$TLI = 1 - \frac{\frac{\chi_I^2}{df_I} - \frac{\chi_t^2}{df_t}}{\frac{\chi_I^2}{df_I} - 1} \qquad (1.11).$$

Finally, the Root Mean Square Error of Approximation RMSEA is given by

$$RMSEA = \sqrt{\frac{\frac{\chi_t^2 - df_t}{N}}{df_t}} \qquad (1.12).$$

Using equations (1.10) to (1.12) and the values in Table 1.8, we can calculate the CFI, TLI and RMSEA separately for the within and between models. The results are in Table 1.9.

Table 1.9 Fit indices for individual and family level models separately		
	Individual level	Family level
	2 factors	**1 factor**
CFI	1.00	0.97
TLI	1.01	0.96
RMSEA	0.00	0.04

The goodness-of-fit indices in Table 1.9 indicate that the within groups model has an excellent fit, and the between groups model has a good fit. There is no need to modify our final two-level factor model.

1.3 DIRECT ESTIMATION OF THE COVARIANCES AT EACH LEVEL: THE MULTIVARIATE MULTILEVEL APPROACH

Goldstein (1987, 1995) suggest to use a multivariate multilevel model to produce a covariance matrix at the different levels, and to input these into a standard SEM program for further analysis. Therefore, for our family data, we use a multivariate multilevel model, with separate levels for the six intelligence tests, the individual children, and the families. We create six dummy variables to indicate the six intelligence scales, and exclude the intercept from the model. Hence, at the lowest level we have

$$Y_{hij} = \pi_{1ij} d_{1ij} + \pi_{2ij} d_{2ij} + \ldots + \pi_{6ij} d_{6ij} \qquad (1.13).$$

At the individual level we have

$$\pi_{pij} = \beta_{pj} + u_{pij} \qquad (1.14).$$

And at the family level (the third level in the multivariate model), we have

$$\beta_{pj} = \gamma_p + u_{pj} \qquad (1.15).$$

By substitution we obtain

$$Y_{hij} = \gamma_1 d_{1ij} + \gamma_2 d_{2ij} + \ldots + \gamma_6 d_{pij} + u_{1ij} + u_{2ij} + \ldots + u_{6ij} + u_{1j} + u_{2j} + \ldots + u_{6j} \qquad (1.16).$$

8//In sum notation, we have:

$$Y_{hij} = \sum_{h=1}^{6} \gamma_h d_{hij} + \sum_{h=1}^{6} u_{hij} + \sum_{h=1}^{6} u_{hj} \qquad (1.17).$$

The model described by equations (1.13) and (1.14), provides us with estimates of the six item means, and of their variances and covariances at the pupil and school level. Since in this application we are mostly interested in the variances and covariances, RML estimation is preferred to FML estimation. Table 1.9 below presents the RML estimates of the covariances and the corresponding correlations at the individual level, and Table 1.10 presents the same at the family level.

Table 1.9 Covariances and correlations at the individual level

	1	**2**	**3**	**4**	**5**	**6**
Word list	9.65	*.24*	*.31*	*.20*	*.08*	*.05*
Cards	3.21	18.38	*.49*	*.14*	*.13*	*.01*
Matrices	3.57	7.89	13.92	*.14*	*.10*	*.06*
Figures	2.59	2.56	2.13	17.93	*.26*	*.19*
Animals	1.03	2.41	1.61	4.74	18.03	*.45*
Occupations	0.55	0.12	0.93	3.19	7.52	15.35

Note: italic entries in upper diagonal are the correlations

Table 1.10 Covariances and correlations at the family level

	1	**2**	**3**	**4**	**5**	**6**
Word list	5.92	*.68*	*.77*	*.20*	*.90*	*.45*
Cards	5.29	10.37	*.81*	*.72*	*.58*	*.25*
Matrices	3.02	4.18	2.57	*.84*	*1.01*	*.59*
Figures	0.89	4.29	2.50	3.42	*.34*	*.07*
Animals	4.91	4.18	3.64	1.43	5.06	*.27*
Occupations	2.77	2.08	0.59	0.35	1.57	6.49

Note: italic entries in upper diagonal are the correlations

Table 1.9 is equivalent to the pooled within families covariance matrix in Table 1.2. The actual values in both tables are certainly very close. Table 1.10, the between families matrix, is *not* equivalent to the scaled between families matrix in Table 1.3. The covariance matrix in Table 1.3 is the scaled between groups matrix, which is equals to the within groups matrix plus the between groups matrix multiplied by the average group size. The average group size for these

family data, indicated by the scaling factor in equation (1.6) is 5.04, and as a result the actual values in Table 1.3 are rather large. Table 1.10 is a direct estimate of the between family covariance matrix itself. It is a maximum likelihood estimator of the population between family covariance matrix, and can be entered directly into a standard SEM program for analysis. As a result, if we are interested in the population structure of the within families or between families covariances, we can input the corresponding sample matrix from Table 1.9 or 1.10 directly into a SEM program. There is no need to analyze them simultaneously, using the two-group option, unless we want to impose constraints across the two levels.

There is one problematic entry in the family level matrix; the correlation between 'matrices' and 'animals' is estimated as 1.01, an impossible value. It is interesting to note that this impossible value is associated with the scale 'matrices', which in the pseudobalanced estimates in Table 1.6 has a small negative variance on the family level. Both impossible values point to the same variable, and the source of the problem is in both cases the small sample size at the family level. In the case of the negative variance estimate, the usual solution is to fix its value to zero. Some SEM programs can continue estimation with an improper input matrix (e.g., Amos, cf. Arbucle & Wothke, 1999). For our data, the result will be a negative variance estimate, which subsequently must be fixed at zero. If a SEM program cannot accommodate an improper covariance matrix, a practical solution is the *ridge* option, which is to multiply the diagonal of the covariance matrix with a number slightly larger than one (the Lisrel program does this automatically, cf. Jöreskog & Sörbom, 1996). In our case, the automatic ridge option in Lisrel multiplies all diagonal values by 2.0 to obtain a proper input matrix.[1]

If we analyze the individual level and family level covariance matrices separately, we find the parameter estimates reported all together in Table 1.11. For the individual level, the chi-square is 7.12 (df=8, p=.52), with all fit indices indicating a good fit. For the family level (after the ridge-correction) the chi-square is 6.47 (df=9, p=.69), with all fit indices indicating a good fit.

[1] The ridge factor chosen by Lisrel is rather large. Lisrel can be instructed to use a smaller value, such as 1.1, which would be sufficient. The Amos estimates on the improper matrix are close to the Lisrel estimates, but for an improper matrix Amos does not calculate the chi-square test. The ridge-solution presented here can be followed using any of the available SEM software.

Table 1.11 Individual and family level estimates, via direct estimation

	Individual level			Family level	
	Numeric	**Perception**	**resid. var.**	**General**	**resid. var.**
Word list	1*		8.15 (1.04)	1*	8.26 (2.26)
Cards	2.23 (.61)		10.91 (2.38)	1.32 (.56)	14.51 (3.96)
Matrices	2.34 (72)		5.64 (2.32)	0.93 (.34)	2.06 (.92)
Figures		.66 (.21)	15.78 (1.99)	0.61 (.30)	5.49 (1.40)
Animals		1.50 (.53)	6.86 (3.86)	1.07 (.42)	6.03 (.1883
Occupations		1*	10.39 (2.07)	$0.34\ (.37)^{ns}$	12.58 (2.99)

Standard errors in parentheses. Correlation between individual factors: 0.21^{ns}; * = fixed; ns = not significant

The estimates in Table 1.6 are all close to the pseudobalanced estimates presented earlier. The residual variances at the family level are very large, but this is the consequence of the large ridge multiplication factor.

The fact that the individual level (within families) and family level (between families) covariances are estimated directly, and consequently can be modeled directly and separately by any SEM program, is a distinct advantage of the multivariate multilevel approach. As a result, we get separate model tests and fit indices at all levels. The multivariate multilevel approach to multilevel SEM also generalizes straightforwardly to more than two levels. There are other advantages as well. First, since the multilevel multivariate model does not assume that we have a complete set of variables for each individual, incomplete data are accommodated without special effort. Second, if we have dichotomous variables, we can use the multilevel generalized linear model to produce the covariance matrices, again without special effort.

There are some disadvantages to the multivariate multilevel approach as well. An important disadvantage is that the covariances produced by the multivariate multilevel approach are themselves estimated values. They are not directly calculated, as the pooled within groups and scaled between groups covariances are, but they are estimates produced by a complex statistical procedure. If the data have a multivariate normal distribution, the pooled within groups and scaled between groups covariances can be viewed as observed values, which have a known sampling distribution. This sampling distribution is used by SEM programs to estimate the chi-square model test and the standard errors of the parameter estimates. It is unknown how well the sampling distribution of the multivariate multilevel covariance estimates follows the sampling distribution of the observed covariances. This is, of course, especially true when we analyze incomplete data or dichotomous variables. In the case of incomplete data it is unclear what the proper sample size is, and in the case of dichotomous variables we know that the underlying distribution is not normal, and that the data in general contain less information than normally

distributed variables do. The covariances estimated using the multivariate multilevel approach are consistent estimates, meaning that as the sample sizes increase they will approach the population values more closely. Most likely, the chi-square model test is upwardly biased, and the standard errors are probably downwardly biased. If the direct approach is used, it seems prudent to use other goodness-of-fit indices in addition to the chi-square test, and to interpret borderline significances with some care.

I will illustrate the advantage of the direct approach when data are incomplete using the Van Peet family data as an example. The data set used so far is just the subset of the data that has no missing values on the six intelligence scales. The full data set does in fact contain many missing values. The main reason for this is that the two last tests 'naming animals' and 'naming occupations' are very time consuming, and for that reason these tests were dropped in the course of the data collection. So, for these two tests, there are 37 families and 187 children, while for the whole (but incomplete) data set there are 49 families and 269 children. Table 1.13. shows the pairwise sample sizes for the individual level correlations. At the family level, there are data on 49 families except for the correlations that involve the two last tests.

Table 1.12 Pairwise sample sizes for family data, individual level

	1	**2**	**3**	**4**	**5**	**6**
Word list	265					
Cards	265	269				
Matrices	265	269	269			
Figures	265	269	269	269		
Animals	187	191	191	191	191	
Occupations	187	191	191	191	191	191

The mean number of individual children is 231 across all entries in the table, and the mean number of families is 43. In this case, using the direct estimation approach on the incomplete data set appears an attractive choice, because we can incorporate far more data in our analysis. Tables 1.13 and 1.14 present the covariances and correlations based on the whole data set. Especially for the family level covariances and correlations, the values for the entire data set appear somewhat different. Since most values in Table 1.14 are based on about 30% more families, the covariances presented here are in all probability more accurate.

Table 1.13 Covariances and correlations at the individual level, entire data set

	1	2	3	4	5	6
Word list	9.07	*.22*	*.30*	*.24*	*.09*	*.06*
Cards	2.82	17.79	*.46*	*.22*	*.15*	*.02*
Matrices	3.36	7.09	13.60	*.18*	*.11*	*.08*
Figures	3.04	3.76	2.71	17.13	*.27*	*.20*
Animals	1.12	2.75	1.67	4.70	18.06	*.46*
Occupations	0.69	0.38	1.12	3.23	7.59	15.30

Note: italic entries in upper diagonal are the correlations

Table 1.14 Covariances and correlations at the family level, entire data set

	1	2	3	4	5	6
Word list	6.30	*.57*	*.81*	*.27*	*.91*	*.52*
Cards	4.12	8.20	*.62*	*.69*	*.46*	*.21*
Matrices	3.61	3.15	3.14	*.66*	*1.02*	*.34*
Figures	1.33	3.91	2.33	3.96	*.37*	*.23*
Animals	5.46	3.15	4.31	1.77	5.74	*.41*
Occupations	3.64	1.73	1.72	1.28	2.75	7.91

Note: italic entries in upper diagonal are the correlations

Given that there is no single number for the sample size, the mean sample size will be used as an indicator for the effective sample size. Thus, the sample size for the family level covariance matrix is 43, and for the individual level covariance matrix 231-43=188. The combined results are in Table 1.15.

Table 1.15 Individual and family level estimates, via direct estimation

	Individual level			Family level	
	Numeric	**Perception**	**resid. var.**	**General**	**resid. var.**
Word list	1*		7.64 (.89)	1*	3.12 (.70)
Cards	2.15 (.56)		11.17 (2.05)	0.87 (.23)	6.15 (1.35)
Matrices	2.29 (65)		6.08 (2.01)	0.95 (.14)	0.01 (.14)
Figures		.67 (.18)	14.81 (1.70)	0.64 (.16)	2.80 (.61)
Animals		1.42 (.41)	7.55 (3.00)	1.19 (.18)	0.91 (.29)
Occupations		1*	10.06 (1.80)	0.48 (.23)	7.83 (1.71)

Standard errors in parentheses. Correlation between individual factors: 0.26^{ns}; * = fixed; ns = not significant

For the individual level, the chi-square is 15.00 (df=8, p=.06), with the fit indices indicating an acceptable fit. For the family level (after the ridge-correction) the chi-square is 75.54 (df=9, p=.00), with all fit indices indicating a poor fit. Earlier we saw that the direct estimation approach on the complete data produces almost the same results as the pseudobalanced approach. The large difference in the family level results when the incomplete cases are added to the analysis, must be the effect of adding extra cases. Given the small sample size at the family level, no further exploration of these data is attempted.

REFERENCES

Arbucle, J.L. & Wothke, W. (1999). *Amos 4.0 User's Guide*. Chicago: SmallWaters.

Bentler, P.M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238-246.

Bentler, P.M. (1995). EQS structural equations program manual. Encino, CA: Multivariate Software.

Bentler, P.M. & Bonett, D.G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.

Bollen, 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Browne, M.W. & Cudeck, R. (1992). Alternative ways of assessing model fit. *Socioloical Methods & Research*, 21, 230-258.

Chou, C.-P., & Bentler, P.M. (1995). Estimates and tests in structural equation modeling. In R.H. Hoyle (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications*. (Pp. 37-55). Newbury Park, CA: Sage

Gerbing, D.W. & Anderson, J.C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. Sociological Methods & Research, 21. 132-161

Goldstein, H. (1987). *Multilevel Models in Educational and Social Research*. London: Griffin.

Goldstein, H. (1995). *Multilevel Statistical Models*. London: Edward Arnold/New York: Halsted.

Goldstein, H. & McDonald, R. (1988). A general model for the analysis of multilevel data. *Psychometrika*, 53, 455-467.

Gustaffson, J.-E., & Stahl, P.E. (1999). *Streams User's Guide, Vs. 2.0*. Mölndal, Sweden: MultivariateWare.
Hays, W.L. (1994). *Statistics*. New York: Harcourt Brace College Publishers.
Heck, R.H. & Thomas, S.L. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Erlbaum.
Hox, J.J. & Maas, C.J.M. (2000). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling*, in press.
Hox, J.J. (1993). Factor analysis of multilevel data: Gauging the Muthén model. In J.H.L. Oud & R.A.W.van Blokland-Vogelesang(Eds.), *Advances in longitudinal and multivariate analysis in the behavioral sciences*, chapter 10 (pp.141-156). Nijmegen: ITS.
Hox, J.J. (1994, revised 2000). *Split2*. Computer Program, available at: http://www.fss.uu.nl/ms/jh.
Hox, J.J. 1995. *Applied multilevel analysis. (Second edition).* Amsterdam: TT-Publikaties. Available at: http://www.fss.uu.nl/ms/jh.
Jöreskog, K.G. & Sörbom, D. (1996). *Lisrel 8: User's Reference Guide. 2nd Edition.* Chicago: Scientific Software International Inc.
Kaplan, D. (1995).Statistical power in SEM. In R.H. Hoyle (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications*. (Pp. 100-117). Newbury Park, CA: Sage
Kaplan, D. & Elliott, P.R. (1997). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling: A Multidisciplinary Journal*, 4, 1-24.
Li, F., Duncan, T.E., Harmer, P., Acock, A., & Stoolmiller, M. (1998). Analyzing measurement models of latent variables through multilevel confirmatory factor analysis and hierarchical linear modeling approaches. *Structural Equation Modeling: A Multidisciplinary Journal.* 5, 3, 294-306.
Longford, N.T. & Muthén, B.O. (1992). Factor analysis for clustered observations. *Psychometrika*, 57, 581-597.
McArdle, J.J. & Hamagami, F. (1996). Multilevel models from a multiple group structural equation perspective. Pp. 89-125 in G.A. Marcoulides & R.E. Schumacker (eds.). *Advanced Structural Equation Modeling*. Mahwah, NJ: Erlbaum.
McDonald, R.P. (1994). The bilevel reticular action model for path analysis with latent variables. *Sociological Methods & Research*, 22, 399-413.
McDonald, R. & Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology*, 42, 215-232.
Muthén, B. & Satorra, A. (1995). Complex sample data in structural equation modeling. Pp. 267-316 in P.V. Marsden (Ed.), *Sociological Methodology, 1995*. Oxford: Blackwell.
Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, 54, 557-585.
Muthén, B. (1990). *Means and Covariance Structure Analysis of Hierarchical Data.* Los Angeles: UCLA Statistics series, #62.
Muthén, B.O. & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. In: Bock, R.D. (ed.). *Multilevel analysis of educational data*. San Diego: Academic Press.
Muthén, B.O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, 22, 376-398.
Muthén, L.K., & Muthén, B.O. (1998). *Mplus. The Comprehensive Modeling Program for Applied Researchers*. Los Angeles: Muthén & Muthén.
Rowe, K.J. & Hill, P.W. (1998). Modeling educational effectiveness in classrooms: the use of multi-level structural equations to model students'progress. *Educational Research and*

Evaluation, 4, 307-347.
Searle, S.R., Casella, G. & McCulloch, C.E. (1992). *Variance Components*. New York: Wiley.
Tucker, C. & C. Lewis (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.
Van Peet, A.A.J. (1992). *De potentieeltheorie van intelligentie*. [The potentiality theory of intelligence] Amsterdam: University of Amsterdam, Ph.D. Thesis.