This article gives an introduction to some new techniques for multilevel covariance structure modeling with latent variables. Although these techniques only incorporate a subset of models that are relevant to multilevel data, the techniques do provide a large set of new analysis possibilities and have the advantage that they only require conventional structural equation modeling software. The presentation draws on methodology presented in earlier works by the author.

Multilevel Covariance Structure Analysis

BENGT O. MUTHÉN
University of California, Los Angeles

This article gives an introduction to some new techniques for multilevel covariance structure modeling with latent variables. Although these techniques only incorporate a subset of models that are relevant to multilevel data, the techniques do provide a large set of new analysis possibilities and have the advantage that they only require conventional structural equation modeling software. The presentation draws on methodology presented in Muthén (1989, 1990). Muthén (1990) provides technical details.

The analysis of multilevel data is a complex topic because it draws on contributions from many different areas of methodological research. Two perspectives can be distinguished, that of sampling and that of varying parameters. From a sampling perspective, multilevel data can be viewed as obtained by cluster sampling. For example, a simple random sample of schools is obtained and within each school a random sample of students is obtained. The analysis needs to specify stochastic variation that mirrors the sampling scheme, such as formulating a model that decomposes the student variation in a school and an individual component. Modeling cluster sampling in this way

AUTHOR'S NOTE: *This research was supported by National Science Foundation grant SES-8821668 and by Grant OERI-G-86-003 from the Office of Educational Research and Improvement, Department of Education. I would like to thank Ginger Nelson Goff, Jin-Wen Yang Hsu, and Kathleen Wisnicki for valuable research assistance.*

reflects the sampling theory tradition of Scott and Smith (1969), Fuller and Battese (1973), Malec and Sedransk (1985), and Battese, Harter, and Fuller (1988). The perspective of varying parameters is more akin to concerns of multilevel writing in the educational literature (Cronbach 1976; Bock 1989), to random coefficient modeling in econometrics (Swamy 1970), and to the repeated measurement literature on individual differences in growth parameters for longitudinal data (Laird and Ware, 1982). Continuing the example of students sampled within schools, the perspective of varying parameters considers a model for relationships among variables observed for the students as having parameters that can obtain different values for different schools.

One can also structure the topic by distinguishing between analyses that estimate the same parameters as would usually be estimated under the conventional assumption of simple random sampling and analyses that also estimate additional parameters due to the multilevel structure. The second type of approach would estimate both student-level and school-level parameters. For a discussion of both type of approaches, see Muthén and Satorra (1991). The present article focuses on the second approach. It is, however, important to be aware of the issues involved in the first approach. Consider again the example of students sampled within schools. Formulating a conventional covariance structure model for the variables observed for these students, the first approach would estimate the usual parameters but use special formulas for computing standard errors of estimates as well as special formulas for computing a chi-square test of model fit. The special formulas are developed without resorting to the incorrect assumption of simple random sampling. The consequences of incorrectly assuming simple random sampling when data are obtained via cluster sampling are customarily studied in terms of design effects (Kish 1965), describing the ratio of the variance of an estimator under cluster sampling to the variance under simple random sampling. For example, from sampling theory it is well-known that the design effect for the sample mean of a random sample of G clusters, each having c elements, is $1 + (c - 1)\rho$, where ρ is the intraclass correlation coefficient measuring the degree of homogeneity of the observations within each cluster (see, e.g., Cochran 1977, p. 262). In our example, the larger the number of students sampled per school (c) and the larger the homogeneity of students with respect to what is measured (ρ), the larger the underestimation of the true variance of the estimator when using a variance expression based on conventional simple random sampling theory. Conventional inference procedures would only be correct if $c = 1$ or $\rho = 0$. The first type of approach would use formulas that give correct inference for the usual set of parameters. For a good overview of relevant issues in survey sampling, see Skinner, Holt, and Smith (1989).

MULTILEVEL COVARIANCE STRUCTURE MODELS

In conventional covariance structure modeling, a p-variate vector $\mathbf{y}_i$ is observed for individual i. For simplicity, we will consider the special case of each $\mathbf{y}_i$ being multivariate normally distributed. The individual observation vectors are assumed to be independently and identically distributed (i.i.d.). Consider this assumption in the setting of students observed within G schools. Table 1 gives an example of the implication of this assumption for the five first individuals of the data matrix. Here, School 1 has three students, School 2 has two students, and so on. The top left of the covariance matrix for all observations is indicated in the right part of Table 1. Due to the assumption of independence, this matrix has zero off-diagonal submatrices whereas the assumption of identically distributed observations states that the diagonal submatrices are identical. Although not shown, the population mean vector is also identical for all observations.

This article focuses on maximum likelihood estimation under normality. Given the Table 1 covariance structure, the likelihood of the sample of all observations can be expressed in terms of the common $p \times p\ \mathbf{\Sigma}$ matrix.

Multilevel covariance structure modeling builds on different assumptions. It relaxes the assumption of identically distributed observations. As a motivating example, we will consider a single-factor model for the students in the set of G schools. To make clear the hierarchical nature of the data, subscripts g and i will be used for schools (group) and students (individual), respectively.

TABLE 1: Conventional Covariance Structure Modeling

School	*Student*	*Data*	*Covariance Structure*				
1	1	y_1	Σ		symmetric		
1	2	y_2	0	Σ			
1	3	y_3	0	0	Σ		
2	1	y_4	0	0	0	Σ	
2	2	y_5	0	0	0	0	Σ

$$\mathbf{y}_{gi} = \boldsymbol{\nu} + \boldsymbol{\lambda}\, \eta_{gi} + \boldsymbol{\epsilon}_{gi} \tag{1}$$

Here, ν is a measurement intercept vector, $\boldsymbol{\lambda}$ is a vector of factor loadings, η represents the factor, and $\boldsymbol{\epsilon}$ represents the residual vector.

As pointed out in Muthén and Satorra (1989), the varying parameter perspective is analogous to random coefficients in regression models. One model is formulated for the individual-level variation and another is formulated for the across-group variation in the parameters of the individual-level model. We might for simplicity assume that only the parameters of factor means vary across groups. In conventional, multiple-group structural equation modeling, this could be interpreted as having G (or G – 1 to be precise) factor means estimated for the G groups (schools). In the multilevel setting, however, schools are viewed as randomly sampled so that instead of fixed parameters, the factor means should be specified by means of random effects. In this way, we may write

$$\eta_{gi} = \alpha + \eta_{Bg} + \eta_{Wgi}, \tag{2}$$

where α is the overall expectation for η_{gi}, η_{Bg} is a random factor component capturing school effects and having zero expectation, and η_{Wgi} is a random factor component varying over students within their respective schools and having zero expectation. Note that conditional on student i being in school g, the mean of the factor η_{gi} is $\alpha + \eta_{Bg}$, where η_{Bg} varies randomly across schools. In this way, only two parameters are needed to capture the school differences in factors, α and the variance of η_{Bg}, ψ_B, say. This random effects specification with two parameters obviously is more parsimonious than the fixed effect, multiple-group specification with $G - 1$ parameters as soon as the number of schools exceeds three. The random effects specification of varying factor means is the natural way to model in the multilevel setting. In this way, the total factor variance may be broken down into a between-school variance component and a within-school variance component

$$V(\eta_{gi}) = \psi_T = \psi_B + \psi_W. \tag{3}$$

From a substantive point of view, it is of interest to estimate the relative size of the between-school factor variation ψ_B relative to the total factor variation ψ_T. Note that using (2) we find that the factor covariance of students within a school is ψ_B, since for two students i and i′,

$$\begin{aligned} \text{Cov}(\eta_{gi}, \eta_{gi'}) &= \text{Cov}(\eta_{Bg}, \eta_{Bg}) + \text{Cov}(\eta_{Wgi}, \eta_{Wgi'}). \\ &= \psi_B \qquad\qquad + 0 \end{aligned} \tag{4}$$

This means that the factor values, and hence the observed scores, are not independent for students who are in the same school. On the factor level, the magnitude of ψ_B describes the strength of the nonindependence. As discussed by Muthén (1991), the latent variable counterpart of an "intraclass correlation" is consequently the ratio

$$\psi_B/(\psi_B + \psi_W). \tag{5}$$

The residual variation of ϵ_{gi} in (1) can also be broken down in a group-level (between) and an individual-level (within) component,

$$V(\epsilon_{gi}) = \Theta_B + \Theta_W. \tag{6}$$

The covariance structure for this random effects model is

$$V(y_{gi}) = \Sigma_T = \Sigma_W + \Sigma_B, \tag{7}$$

where Σ_B is the "between" matrix representing across-school variation,

$$\Sigma_B = \lambda\, \psi_B\, \lambda' + \Theta_B \tag{8}$$

and Σ_W is the within matrix representing within-school, student-level variation,

$$\Sigma_W = \lambda \psi_W \lambda' + \Theta_W. \tag{9}$$

TABLE 2: Multilevel Covariance Structure Modeling

School	*Student*	*Data*	*Covariance Structure*				
1	1	y_{11}	$\Sigma_W + \Sigma_B$	symmetric			
1	2	y_{12}	Σ_B	$\Sigma_W + \Sigma_B$			
1	3	y_{13}	Σ_B	Σ_B	$\Sigma_W + \Sigma_B$		
2	1	y_1	0	0	0	$\Sigma_W + \Sigma_B$	
2	2	y_2	0	0	0	Σ_B	$\Sigma_W + \Sigma_B$

In line with (4), the covariance matrix for students who are in the same school is Σ_B. We now have all the components needed to formulate the multilevel counterpart to the covariance structure for the data matrix of Table 1. Table 2 gives the structure for the same five students as in Table 1. Table 2 reflects the fact that we no longer have independence across all students in that Σ_B appears in certain off-diagonal positions. The diagonal matrices are still constant across all students, but consist of the sum of Σ_W and Σ_B. The multilevel covariance structure of Table 2 is clearly less restrictive than that of Table 1. The conventional modeling of Table 1 may be viewed as restricting Σ_B to zero, assuming zero intraclass correlations. The multilevel modeling also has the advantage of being able to disentangle the variation within and between groups. The separate estimation of Σ_W and Σ_B may be of great substantive interest, as will be shown in the example section.

In some applications, however, only the total variation of Σ_T may be of interest. In this case, the multilevel modeling still provides a Σ_T estimate via the sum of Σ_W and Σ_B and does provide correct inference. As mentioned earlier, simpler approaches are, however, possible in that the parameters of Σ_T can be estimated directly and correct inference can be provided by special calculations of standard errors and chi-square (see Muthén and Satorra 1991).

The covariance structure of Table 2 is more complex than that of Table 1. Whereas the Table 1 structure enables the formulation of a likelihood in terms of a single $p \times p$ Σ matrix, the multilevel likelihood for Table 2 is more complicated. As shown in McDonald and Goldstein (1989) and Muthén (1989, 1990), however, the likelihood can fortunately also here be expressed in terms of $p \times p$ covariance matrices. Muthén (1989, 1990) shows that the likelihood can, in fact, be simply expressed in terms of two covariance matrices, Σ_W and $\Sigma_W + c\ \Sigma_B$, where c reflects the group size.

The multilevel factor modeling discussed above leads to a covariance structure model for two-level data, which uses a conventional factor analysis covariance structure on both the between and within level. As opposed to the covariance structures of (8) and (9), a more general formulation allows the factor loading matrices to differ on the within and between levels so that with multiple factors,

$$y_{gi} = \nu + \Lambda_B \eta_{B_g} + \epsilon_{B_g} + \Lambda_W \eta_{W_{gi}} + \epsilon_{W_{gi}} \qquad (10)$$

$$V(y_{gi}) = \Sigma_B + \Sigma_W, \qquad (11)$$

$$\Sigma_B = \Lambda_B \Psi_B \Lambda_B' + \Theta_B, \text{ and} \qquad (12)$$

$$\Sigma_W = \Lambda_W \Psi_W \Lambda_W' + \Theta_W. \qquad (13)$$

More general structural equation models can also be formulated as in Schmidt and Wisenbaker (1986), McDonald and Goldstein (1989), and Muthén (1989, 1990).

MULTILEVEL COVARIANCE STRUCTURE ESTIMATION

In the two-level case with G groups to be considered here, the likelihood is formulated for G multivariate normal observation vectors, where each vector contains all variables for all individuals in the group. There are N_g individuals in group g, where $N = \Sigma N_g$ is the total sample size. Unlike conventional analysis, independence of observations is not assumed over all *N* observations but only over the G groups, while the intraclass correlation is modeled via Σ_B. The covariance matrices of Σ_B and Σ_W contain the parameters of interest. In this article, we will assume that we study the common case of no mean structure. As opposed to conventional covariance structure analysis, we do not use only the regular $p \times p$ sample covariance matrix. In the balanced case with no mean structure, the customary between and pooled-within sample covariance matrices (see below) provide sufficient information for maximum-likelihood estimation, while the unbalanced case also needs information on each group's mean vector. With maximum-likelihood estimation, a large-sample chi-square variable is obtained to test restrictions imposed by the model on Σ_B and Σ_W. With p variables and r parameters, the number of degrees of freedom is $p(p + 1) - r$. We note that a conventional covariance structure model has $p(p + 1)/2 - r$ degrees of freedom since this analysis restricts the matrix Σ_B to be zero (in this case r is reduced by the number of parameters for the between part). For more details and the relationship to conventional structural equation modeling, see Muthén (1989, 1990).

Although in principal, special formulas and software could be developed for multilevel covariance structure analysis (MCA) maximum-likelihood (ML) estimation, Muthén (1989, 1990) showed that multiple-group structural equation modeling software can be modified for MCA ML analysis. In line with this idea, Muthén proposed a simpler ML-based MCA estimator, which can be used with already existing multiple-group structural equation software such as LISREL, LISCOMP, and EQS. This estimator uses the customary between and pooled-within sample covariance matrices. In the balanced case, it is equivalent to the MCA ML estimator. In the unbalanced case, the estimator is consistent and, despite the fact that it uses less information than ML, has given similar results in the analyses to date (Muthén 1990). The true ML procedure will be referred to as FIML (full information ML) and the simpler estimator as MUML (Muthén's ML-based estimator). We will use both procedures in our analyses for comparison purposes.

The MUML estimator of Muthén (1989, 1990) demonstrates the basic features of MCA. Consider the three customary sample covariance matrices S_T, S_{PW}, S_B,

$$S_T = (N - 1)^{-1} \sum_{g=1}^{G} \sum_{i=1}^{N_g} (y_{gi} - \bar{y})(y_{gi} - \bar{y})' \qquad (14)$$

$$S_{PW} = (N - G)^{-1} \sum_{g=1}^{G} \sum_{i=1}^{N_g} (y_{gi} - \bar{y}_g)(y_{gi} - \bar{y}_g)' \qquad (15)$$

$$S_B = (G-1)^{-1}\sum_{g=1}^{G} N_g (\bar{y}_g - \bar{y})(\bar{y}_g - \bar{y})' \tag{16}$$

The matrix S_T is used in conventional covariance structure analysis. In the multilevel case, it is a consistent estimator of the total covariance matrix $\Sigma_B + \Sigma_W$. In line with expected mean squares developments in univariate analysis of variance (cf. Winer, Brown, and Michels 1991), one can show that the pooled-within matrix S_{PW} is a consistent and unbiased estimator of Σ_W, while the between matrix S_B is a consistent and unbiased estimator of

$$\Sigma_W + c\,\Sigma_B, \tag{17}$$

where c reflects the group size,

$$c = \left[N^2 - \sum_{g=1}^{G} N_g^2\right][N(G-1)]^{-1}. \tag{18}$$

For balanced data, c is the common group size. For unbalanced data and large number of groups, c is close to the mean of the group sizes. Note that the between matrix S_B is the covariance matrix of group means $\bar{y}_g$ weighted by the group size. Equation (17) shows that the population counterpart of S_B is a function of both Σ_B and Σ_W. The ML estimate of Σ_W is S_{PW}, while the ML estimate of Σ_B is (Muthén 1990)

$$c^{-1}(S_B - S_{PW}). \tag{19}$$

the MUML estimator (Muthén 1989, 1990) minimize the fitting function,

$$G\{\ln|\Sigma_W + c\Sigma_B| + \mathrm{trace}[(\Sigma_W + c\Sigma_B)^{-1}S_B] - \ln|S_B| - p\} + (N-G)\{\ln|\Sigma_W| + \mathrm{trace}[\Sigma_W^{-1}S_{PW}] - \ln|S_{PW}| - p\}. \tag{20}$$

Here, G is the number of groups, c is defined in (18), p is the number of variables, N is the total number of observations, and S_B and S_{PW} are the conventional between and pooled-within sample covariance matrices of (15) and (16). This fitting function is analogous to that of a conventional two-population (two-group) covariance structure analysis using ML estimation under normality. A sample of G observations is considered for the first population while for the second, N – G observations are used. The S_B and S_{PW} sample matrices are used to fit their corresponding population quantities. This implies that the MUML estimation of multilevel factor analysis parameters can be performed by the ML fitting function in conventional multiple-group structural equation software. The chi-square and standard errors given by the software are rough approximations to the correct values. The quasi-chi-square test of model fit refers to the testing of H_0 against unrestricted Σ_B and Σ_W matrices as is desired. The S_B and S_{PW} matrices can be obtained via standard statistical packages. The author has written a program, available to anyone who wants it, which computes these two matrices, the c value, the intraclass correlations, and the ML estimate of Σ_B for two-level data. Instructions for arranging the data to be able to use this program are given in Nelson and Muthén (1991). This means that the MUML estimator is easily accessible today, while this is not true, in general cases, for FIML. The FIML estimator uses a fitting function similar to (20), but involves terms for each distinct group size, including information on the mean vectors (Muthén 1990). Even when FIML can be done, it will be computationally heavier than MUML as the number of distinct group sizes increases. It should be noted that problems of nonconvergence due to a poor choice of starting values appear to be more common in multilevel factor analysis (MFA) than in conventional covariance structure modeling. The next section specifies a series of analysis steps, which make for a more informed choice of starting values.

MULTILEVEL COVARIANCE STRUCTURE ANALYSIS PATH DIAGRAMS AND SOFTWARE IMPLEMENTATION

Muthén (1990) showed that the input specification for the structural equation modeling software needed for MUML using (20) can be conveniently indicated via conventional path diagrams. Using a one-factor model for both between and within leads to the model diagram of Figure 1. This diagram follows the notation of (10). Below the row of squares are variables on the within level, ε_W and η_W. This part of the diagram corresponds to a conventional one-factor model. Above the row of squares is a row of circles corresponding to the between components of the observed variables, denoted y_B. In this way, the

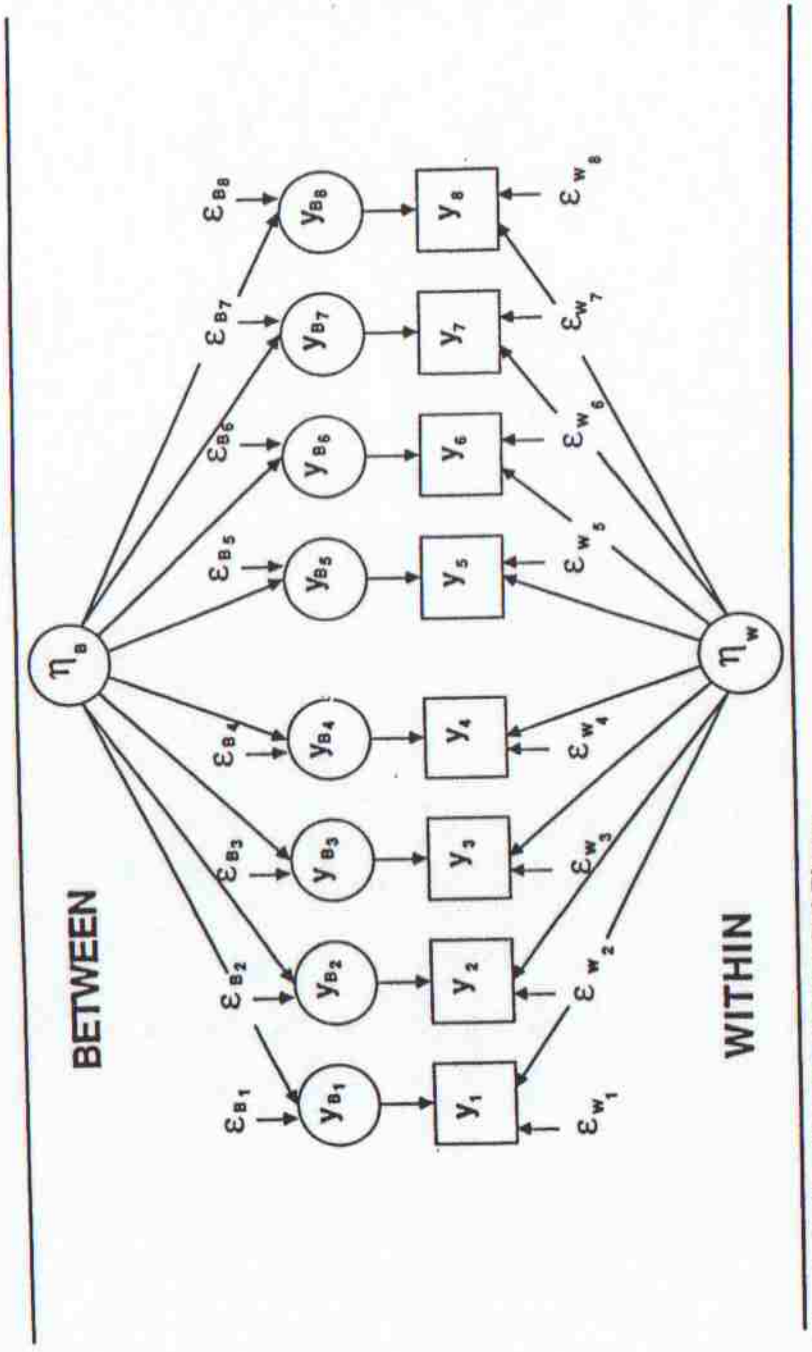

Figure 1: Multilevel Covariance Structure Path Diagram

observed variables y in the squares are functions of within and between components. The between components follow a one-factor model with residuals ε_B and factor η_B.

The path diagram corresponds directly to the first group in the two-group setup indicated by (20). The first group involves the covariance matrix structure $\Sigma_W + c\Sigma_B$. This deviates from the total covariance matrix $\Sigma_W + \Sigma_B$ by the scalar multiplier c for the between part. This means that the between components of the variables have to be scaled by $\sqrt{c}$, which is accomplished by letting the paths (loadings) from the y_Bs to the ys have coefficients $\sqrt{c}$. The second group in (20) corresponds to the within variation. The covariance structure of Σ_W is captured by using the same model structure as for the first group, following Figure 1, but fixing all between coefficients and variance-covariance parameters to zero. Because Σ_W also appears in the covariance structure of the first group, equality restrictions across groups need to be applied for the within parameters.

STRATEGIES FOR MULTILEVEL COVARIANCE STRUCTURE ANALYSIS

As pointed out in Muthén (1989), MCA is a complex analysis, which needs to follow a sound strategy. The actual MCA should, in a typical case, be preceded by four important analysis steps: conventional factor analysis of S_T, estimation of between variation, estimation of within structure, and estimation of between structure.

Step 1: Conventional factor analysis of S_T. This analysis is useful to try out model ideas. The analysis is incorrect when the data is multilevel due to the correlated observations. The model test of fit is usually inflated, particularly for data with large intraclass correlations, large class sizes, and highly correlated variables. However, the test of fit might still be of practical use by giving a rough sense of fit.

Step 2: Estimation of between variation. It is wise to first check if a multilevel analysis is warranted by testing $\Sigma_B = 0$. This can be carried out in an MCA. A simpler way, however, to get a rough indication of the amount of between variation is to compute the estimated intraclass correlations for each variable obtained as the estimate of

$$\sigma_B^2/(\sigma_B^2 + \sigma_W^2). \quad (21)$$

In line with (19), σ_W^2 is estimated as s_{PW}^2 and σ_B^2 is estimated as

$$c^{-1}(s_B^2 - s_{PW}^2). \quad (22)$$

These estimates may be obtained by random effects ANOVA (Winer, Brown, and Michels 1991). The author's program, mentioned above, can also be used. If all intraclass correlations are close to zero, as is the case for many applications, it might not be worthwhile to go further. A good overview of intraclass correlation estimation is given in Koch (1983).

Step 3: Estimation of within structure. If the multilevel model is correct, a conventional covariance structure analysis of S_{PW} is the same as an MCA with an unrestricted Σ_B matrix. This analysis estimates individual-level parameters only. Experience has shown that the analysis gives estimates that are close to the within parameters of an MCA. The conventional analysis would use a sample size of $N - G$ and either the normal theory GLS or ML estimator. Since the S_{PW} analysis is not distorted by the between covariation, it is expected to give a better model fit than the S_T analysis (see also Keesling and Wiley 1974; Muthén 1989) and it is, therefore, the preferred way to explore the individual-level variation.

Step 4: Estimation of between structure. The analysis of between structure is the more difficult part of multilevel analysis. Little might be known about the covariance structure of Σ_B because it does not concern the customary individual-level data but instead across-group (co)variation. The between components have a different meaning than the within components, and it is not clear that the between-group covariation follows a simple model. As analyses in Cronbach (1976) and Harnqvist (1978) have shown, the same structure as that seen in the within level cannot be expected. It is tempting to use S_B to explore the between structure. Note, however, that S_B is not an unbiased or consistent estimator of Σ_B as is indicated in (17). The Σ_B estimator is also a function of S_{PW}. In other words, any simple structure expected to hold for Σ_B does not necessarily hold for S_B, but it should hold within

sampling error for the ML estimate of Σ_B. Unfortunately, the ML estimator of Σ_B is frequently not positive definite and might not even have positive variance estimates. This means that, in practice, we might have to resort to analyzing S_B to get a notion of the Σ_B structure. Fortunately, experience shows that when it is possible to analyze both matrices, similar results are obtained. It might be noted that the ML option of conventional software gives very distorted chi-square test of fit values when using the estimated Σ_B matrix. An alternative is to use MCA with an unrestricted Σ_W matrix (see also Longford and Muthén 1992), only testing the restrictions on Σ_B. One might also obtain an MCA estimate of Σ_B using the Σ_W structure indicated in Step 3 and submit this estimate to covariance structure analysis.

The next set of steps uses the outcomes of the four initial steps to specify a sequence of MCAs. As is shown in (20), the MCA makes use of S_{PW} and S_B simultaneously. The computations are not complicated by a nonpositive definite Σ_B estimate, since this matrix only appears in the sum $\Sigma_W + c\Sigma_B$.

AN EXAMPLE

In the Second International Mathematics Study (SIMS; Crosswhite, Dossey, Swafford, McKnight, and Cooney 1985), a national probability sample of school districts was selected proportional to size; a probability sample of schools was selected proportional to size within school district; and two classes were randomly drawn within each school. We will consider a subset of the U.S. eighth-grade data of 3,724 students who took the core test at both the pretest in fall of 1982 and posttest in spring of 1983. These students were observed in 197 classes from 113 schools. The class sizes vary from 2 to 38, with a typical value of around 20.

The core test consisted of 40 items in the areas of arithmetic, algebra, geometry, and measurement. The topics covered in these items were broken down into eight subscores, where each subscore is the sum of binary items. The subscore RPP consists of eight ratio, proportion, and percentage items. FRACT consists of eight common and decimal fraction items. EQEXP consists of six algebra items involving equalities and expression. INTNUM consists of two items involving integer number algebra manipulations. STESTI consists of five items dealing with measurement items involving standard units and estimation. AREAVOL consists of two measurement items dealing with area and volume determination. COORVIS consists of three geometry items involving coordinates and spatial visualization. PFIGURE consists of five geometry items involving properties of plane figures.

The analysis strategy suggested in the previous section will now be applied to the eight achievement variables at both pretest and posttest. A two-level MFA for students within classes will be used in all cases. The school level will be ignored here for simplicity. Because there are only two classes per school and the school variance proportions are relatively small (Muthén 1991), this clustering effect should not seriously bias the results. A single-factor model is expected to hold reasonably well on the individual level, given that mathematics skills are rather undifferentiated in the eighth grade and are likely to reflect a single, general dimension.

Until now, factor analyses of educational data have routinely ignored the multilevel character of the data. Because of this, it is of interest to compare the results of conventional factor analyses with those of MFA. Various MFA approaches will also be compared. In this way, the Step 1 analysis of S_T will be contrasted with MFA, and in terms of MFA, the traditional estimation via S_{PW} and S_B used in Steps 3 and 4 will be contrasted with MUML and FIML.

INITIAL ANALYSES IN FOUR STEPS

Step 1: Conventional factor analysis of S_T. Table 3 shows the one-factor chi-square tests of model fit and estimated item characteristics for pretest, whereas Table 4 gives the same values for posttest. Standard errors of estimates will not be given in this article because all models presented show parameters significantly different from zero due to the large sample size. The univariate skewness and kurtosis values in these tables do not indicate substantial deviations from the assumed normality, which might have been the case given the small number of items forming the subscores. The S_T analysis gives a reasonable fit at both pretest and posttest, given the large sample size

TABLE 3: Pretest Factor Analysis Results

	Model tests	
Method	Chi-Square[a]	df
S_T	83.71	20
S_{PW}	58.29	20
MFA		
MUML	106.16	40
FIML	98.91	40

	Item Characteristics							
				Reliability				
						MFA		
MFA	Skewness	Kurtosis	Proportion Between	S_T	S_{PW}	Within	Between	MFA Error-Free Proportion Between
RPP	.38	–.68	.34	.61	.44	.44	.96	.52
FRACT	.37	–.57	.39	.60	.38	.38	.97	.61
EQEXP	.23	–.57	.27	.36	.18	.18	.83	.64
INTNUM	.60	–.80	.27	.34	.18	.18	.81	.63
STESTI	–.24	–.64	.32	.44	.25	.25	.86	.61
AREAVOL	.68	–.89	.18	.29	.18	.18	.82	.50
COORVIS	.38	–.70	.21	.34	.18	.18	.92	.57
PFIGURE	.61	–.21	.24	.32	.17	.17	.78	.59

NOTE: MFA = multilevel factor analysis; MUML = Muthén's maximum-likelihood-based estimator; FIML = full information ML; RPP = eight ratio, proportion, and percentage items; FRACT = eight common and decimal fraction items; EQEXP = six algebra items involving equalities and expression; INTNUM = two items involving integer number algebra manipulations; STESTI = five items dealing with measurement items involving standard units and estimation; AREAVOL = two measurement items dealing with area and volume determination; COORVIS = three geometry items involving coordinates and spatial visualization; PFIGURE = five geometry items involving properties of plane figures.

a. For MUML, A quasi chi-square value is given.

TABLE 4: Posttest Factor Analysis Results[a]

Method	Model tests: Chi-Square[b]	df
S_T	88.59	20
S_{PW}	57.45	20
MFA		
MUML	116.00	40
FIML	128.89	40

MFA	Item Characteristics: Skewness	Kurtosis	Proportion Between	Reliability: S_T	Reliability: S_{PW}	Reliability: MFA Within	Reliability: MFA Between	MFA Error-Free Proportion Between
RPP	0.03	−1.07	.38	.68	.52	.52	.97	.53
FRACT	−0.01	−0.92	.40	.68	.49	.49	.98	.57
EQEXP	−0.02	−0.89	.38	.55	.32	.32	.92	.64
INTNUM	−0.07	−1.41	.30	.43	.25	.25	.88	.61
STESTI	−0.44	−0.62	.33	.52	.34	.34	.89	.56
AREAVOL	0.16	−1.44	.25	.38	.23	.23	.84	.54
COORVIS	−0.03	−1.00	.30	.42	.26	.26	.80	.55
PFIGURE	0.15	−0.94	.33	.46	.31	.31	.77	.54

a. See Table 2 for definitions
b. For MUML A quasi chi-square value is given.

of 3,724. This sample size makes the power of the test high and thus rejection at the 5% level might reflect trivial deviations from the model. We note again that this test is incorrect, given that the hierarchical nature of the data is ignored. The chi-square value is most likely inflated.

Step 2: Estimation of between variation. The proportion between variation, or intraclass correlation, for the eight items are in the range .18-.39 at pretest and .24-.40 at posttest. The values increase over time for all variables and particularly for EQEXP and PFIGURE. Note that individual-level measurement error contributes to the within variances. Due to this, individual level measurement error probably deflates the intraclass correlations. The fact that they are still large makes it reasonable to proceed to Step 3.

Step 3: Estimation of within structure. The third step carries out the analysis of the pooled-within matrix $\mathbf{S}_{PW}$. For both pretest and posttest, the conventional ML analysis gives a worse fit for $\mathbf{S}_T$ than $\mathbf{S}_{PW}$ for the one-factor model. The difference in number of observations is negligible, $N = 3{,}724$ versus $N - G = 3{,}527$ and cannot alone explain the difference. The worsening of fit is expected, given the large size of the intraclass correlations and the large average class size of about 20. Judging from the $\mathbf{S}_{PW}$ analysis, the within part of the model has a very good fit to the one-factor model, given the large sample size. It is also interesting to note from Tables 3 and 4 that, relative to the $\mathbf{S}_{PW}$ analysis, the conventional analysis of $\mathbf{S}_T$ strongly overestimates the reliabilities of the variables as estimated by the factor model. The $\mathbf{S}_{PW}$ analysis adjusts for differences in class means. Heterogeneity in the means across classes increases the reliable part of the variation, which inflates the reliabilities (see, also, Muthén 1989, pp. 559-60). The $\mathbf{S}_T$ reliabilities might be correct for inference to this particular mixture of class means, but is not correct for the student scores in any of the classes. This is further discussed below in connection with the MFA results. It appears that the conventional $\mathbf{S}_T$ factor analysis of students sampled within classes can be quite misleading.

Step 4: Estimation of between structure. In the fourth step, we investigate the between structure. The estimated Σ_B was scaled to a correlation matrix and subjected to ordinary exploratory factor analysis by unweighted least squares. Judging from the eigenvalues, a one-factor model holds at both pretest and posttest. For pretest, the first four values are 7.08, 0.26, 0.21, 0.17, whereas for posttest, they are 6.79, 0.30, 0.25, 0.21. The two-factor solutions had no interpretable structure. The analysis of the correlation matrix corresponding to $\mathbf{S}_B$ gave similar results. The estimated loadings are rather close to those obtained via the estimated Σ_B, although somewhat lower overall. MFA analysis using an unrestricted Σ_W matrix and a one-factor model for Σ_B results in a 20 degree of freedom MUML quasi-chi-square value of about 50 for the pretest and posttest, indicating a reasonable fit for the one-factor between structure. In passing, we may note that the corresponding MUML test of fit for the model with an unrestricted Σ_B and a one-factor model for Σ_W resulted in slightly higher quasi-chi-square values with the same degrees of freedom. These values correspond to the chi-square values given for $\mathbf{S}_{PW}$ in Tables 3 and 4. In this sense, the between structure has no worse fit than the within structure.

MFA ANALYSES

The four initial analysis steps suggest an MFA model with one factor for both within and between. This is the model of Figure 1. As in conventional factor analysis, the metric of each factor has to be determined, and this is done by fixing the between and within loadings for RPP to unity.

The MUML quasi-chi-square tests of model fit are 106.16 and 116.00 for pretest and posttest with 40 degrees of freedom. Given the sample size of 3,724, this is taken as a good fit. As is frequently observed, the corresponding proper chi-square values of FIML are rather close, 98.91 and 128.89. Given the unusually large range of class sizes (2-38) in this application, the data are far from balanced and the MUML approximation to FIML is put to a hard test. MUML estimates are in no case more than 7% off from the FIML estimates (not reported) and are usually much closer. The item characteristics deduced from these estimates in Tables 3 and 4 typically differ by 0.01 between MUML and FIML. The within values are almost exactly the same.

The $\mathbf{S}_{PW}$ analysis fitted the within part of the model with 20 degrees of freedom, which may be viewed as an analysis with no between

structure imposed. The addition of the between structure in the MFA adds about 50 to 60 chi-square points, for an additional 20 degrees of freedom. This increase does not seem unduly large for the sample size. It is interesting to note the perfect agreement, to two digits, in the estimated within reliabilities for MFA and $\mathbf{S}_{PW}$. This is because the MFA estimation of Σ_W is largely determined by the second group in the MUML fitting function of (20) due to the large number of students per class.

The MFA within reliabilities are very low, as is expected, given the small number of items comprising each subscore. There is a strong increase over time, particularly for EQEXP and PFIGURE. These correspond to new topics at pretest for many eighth graders, whereas they have been better covered at posttest.

The $\mathbf{S}_{PW}$ analysis gives reliabilities that agree with the within values of the MFA to two digits. The higher $\mathbf{S}_T$ values observed above might be viewed in terms of the MFA model. For simplicity, assume that to a reasonable approximation, λ_{Bj} equals λ_{Wj}. Then the reliable part of the $\mathbf{S}_T$ variance is modeled as $\lambda_j^2(\sigma^2_{\eta B} + \sigma^2_{\eta W})$ while the error variance sums the between and within errors. The reliable part thereby increases which, taken together with a relatively small between error variance, results in the $\mathbf{S}_T$ reliability overestimation. In this application, both the pretest and posttest data led to a rejection of the test of equality of between and within loadings. This might be due to the large sample size, however, because the pattern of estimated loadings is very similar.

The between reliabilities are very high and the indicators of the between factor are very similar. It might be noted that the Step 4 one-factor analysis of the estimated Σ_B gave between reliabilities that are almost identical to the MFA results. Step 4 analysis based on $\mathbf{S}_B$, however, gives consistently lower between reliabilities. The MFA estimation with unrestricted Σ_W and a one-factor Σ_B resulted in values close to those obtained by using the estimated Σ_B, as should be the case because the within structure fits rather well.

The right-most columns of Tables 3 and 4 give the intraclass correlation of the factors, defined as the true intraclass correlation by the ratio ψ_B/ψ_T given in (5). These values are around 0.6 and do not change much from pretest to posttest. These intraclass correlation values might be compared to the observed variable counterparts under the heading "Proportion between." The latter values range from 0.2 to 0.4, with higher values at posttest. In this way, one method can be taken to say that between-class variation does not increase relative to within-class variation across eighth grade, whereas the other says that it does. The substantive implications in terms of effects of tracking students into classes with different curricula are quite different. The latter values are, however, attenuated due to individual-level measurement error, which inflates the within variance part. The attenuation is stronger at pretest than at posttest resulting from the fact that measurement error decreases across time as a function of an increase in exposure to new topics. Distorted comparisons of intraclass correlations across time are thereby obtained. In contrast, MFA takes measurement error into account, avoids the underestimation of true between-class variation, and avoids the distorted across-time comparison of intraclass correlations (see also Muthén 1991).

A final methodological note is of interest regarding the influence of the between structure on the results. As has been pointed out, the between structure might be difficult to determine or might not be a simple one. If the research interest is not in the between structure per se, but only in correctly accounting for the between (co)variation, a simpler alternative approach exists if it can be assumed that the between error variation is negligible. This approach consists of using an MFA with an unrestricted Σ_B. Such a model avoids committing to a specific between structure. Using the assumption of the between error being negligible relative to the within error of this model still enables the calculation of error-free between variance proportions, as in Tables 3 and 4. Here, ψ_B of (5) is replaced by the estimated Σ_B variances. The usually larger within error is still taken into account. Applying this approach to the posttest gave error-free between proportion values that overestimated those of Table 4 by no more than 0.07 for the last two variables having the lowest between reliabilities and no more than 0.03 for the first five variables having the highest between reliabilities.

Further applications of the MUML approach are given in Gold and Muthén (forthcoming) and Harnqvist, Gustafsson, Muthén, and Nelson (forthcoming). As shown in Muthén (1990), the estimation approach can also include group-level variables. For an application using classroom-level information on opportunity to learn, see Muthén (1990). Furthermore, the approach can be directly extended to more than two levels of hierarchical data using more than one between-group structure.

REFERENCES

Battese, G. E., R. M. Harter, and W. A. Fuller. 1988. "An Error-Components Model for Prediction of County Crop Areas Using Survey Satellite Data." *Journal of the American Statistical Association* 83:23-36.

Bock, R. D. 1989. *Multilevel Analysis of Educational Data*. San Diego, CA: Academic Press.

Cochran, W. G. 1977. *Sampling Techniques*. 3rd ed. Toronto: Wiley.

Cronbach, L. J. 1976. "Research on Classrooms and Schools: Formulation of Questions, Design, and Analysis." Unpublished manuscript, Stanford University, Stanford Evaluation Consortium, School of Education.

Crosswhite, F. J., J. A. Dossey, J. O. Swafford, C. C. McKnight, and T. J. Cooney. 1985. *Second International Mathematics Study: Summary Report for the United States*. Champaign, IL: Stipes.

Fuller, W. A. and G. E. Battese. 1973. "Transformations for Estimation of Linear Models With Nested-Error Structure." *Journal of the American Statistical Association* 68:626-32.

Gold, K., and B. Muthén, B. Forthcoming. "Extensions of Covariance Structure Analysis: Hierarchical Modeling of Multidimensional Achievement Data." *Journal of Educational Measurement*.

Harnqvist, K. 1978. "Primary Mental Abilities of Collective and Individual Levels." *Journal of Educational Psychology* 70:706-16.

Harnqvist, K., J.-E. Gustafsson, B. Muthén, and G. Nelson. Forthcoming. "Hierarchical Models of Ability at Individual and Class Levels." *Intelligence*.

Keesling, J. W. and D. E. Wiley. 1974. "Regression Models of Hierarchical Data." Paper presented at the Annual Meeting of the Psychometric Society, Palo Alto, CA, March.

Kish, L. 1965. *Survey Sampling*. New York: Wiley.

Koch, G. G. 1983. "Intraclass Correlation Coefficient." *Encyclopedia of Statistical Sciences* 4:212-17.

Laird, N. M. and J. H. Ware. 1982. "Random-Effects Models for Longitudinal Data." *Biometrics* 65:581-90.

Longford, N. T. and B. Muthén. 1992. "Factor Analysis for Clustered Observations." *Psychometrika* 5:581-597.

Malec, D. and J. Sedransk. 1985. "Bayesian Inference for Finite Population Parameters in Multistage Cluster Sampling." *Journal of the American Statistical Association* 80:897-40.

McDonald, R. P. and H. Goldstein. 1989. "Balanced Versus Unbalanced Designs for Linear Structural Relations in Two-Level Data." *British Journal of Mathematical and Statistical Psychology* 42:215-32.

Muthén, B. 1989. "Latent Variable Modeling in Heterogeneous Populations. Presidential Address to the Psychometric Society, July 1989." *Psychometrika* 54:557-85.

———. 1990. "Mean and Covariance Structure Analysis of Hierarchical Data." Paper presented at the Psychometric Society meeting in Princeton, New Jersey, June.

———. 1991. "Multilevel Factor Analysis of Class and Student Achievement Components." *Journal of Educational Measurement* 28:338-54.

Muthén, B. and A. Satorra. 1989. Multilevel aspects of varying parameters in structural models. Pp. 87-99 in *Multilevel Analysis of Educational Data*, edited by R. D. Bock. San Diego: Academic Press.

———. 1991. *Complex Sample Data in Structural Equation Modeling*. Manuscript submitted for publication.

Nelson, G. and B. Muthén. 1991. *Analysis Preparation Steps for Multilevel Analysis Using LISCOMP* (Technical Report). Los Angeles, University of California.

Schmidt, W. and J. Wisenbaker. 1986. *Hierarchical Data Analysis: An Approach Based on Structural Equations*. CEPSE, No. 4., Research Series. East Lansing, MI: Department of Counseling Educational Psychology and Special Education.

Scott, A. and T.M.F. Smith. 1969. "Estimation in Multi-Stage Surveys." *Journal of the American Statistical Association* 64:830-40.

Skinner, C. J., D. Holt, and T.M.F. Smith. 1989. *Analysis of Complex Surveys*. Chichester: Wiley.

Swamy, P.A.V.B. 1970. "Efficient Inference in a Random Coefficient Regression Model." *Econometrica* 38:311-23.

Winer, B. J., D. R. Brown, and K. M. Michels. 1991. *Statistical Principles in Experimental Design*. New York: McGraw-Hill.

Bengt O. Muthén obtained his Ph.D. in statistics at the University of Uppsala, Sweden, and is currently a professor at the Graduate School of Education at the University of California, Los Angeles. His research interests include latent variable modeling, especially in the area of categorical and other nonnormal data. He developed the LISCOMP computer program, which implemented many of his statistical procedures in a structural equation modeling framework. He was the 1988-89 president of the Psychometric Society.

A two-level (hierarchical) model for path analysis with latent variables is described, together with some properties of a computer program written to implement the model. A simple illustrative example is given.

The Bilevel Reticular Action Model for Path Analysis With Latent Variables

RODERICK P. McDONALD
University of Illinois

The object of this article is to give a relatively nontechnical account of a general model given by McDonald and Goldstein (1988, 1989), illustrated by results from a computer package under development for the application of that model. As other contributions to this issue will make clear, in the recent spate of activity on the construction of suitable statistical models for multilevel data, the main line of development has been concerned with the regression of a single response (dependent) variable on one or more fixed (explanatory) variables, expressed as a random slopes and random intercepts (variance components) model (see Kreft, Kim, and de Leeuw 1990, for a review of developments in theory and computer programs). For convenience, we will hereafter refer to this class of models as fixed-independent-single-response models. It is not easy to generalize such random-coefficients models to yield counterpart structural models—path analysis with latent variables—for multivariate data, essentially because in structural models, generally, all variables are random. de Leeuw (1985) has shown that theory for a fixed-independent- single-response model can, in principle, be applied with little modification to fit a multilevel recursive path model with random exogenous and endogenous variables, provided that the random path coefficients of distinct variables are mutually independent. Goldstein (1986) pointed out that an h-level fixed-independent-single-response model could be applied to give an $(h-1)$-level model for a

SOCIOLOGICAL METHODS & RESEARCH, Vol. 22, No. 3, February 1994 399-413