Title
Latent Variable Modeling in Heterogeneous Populations

Author
Bengt O. Muthén

Publication Date
2011-10-24

PSYCHOMETRIKA—VOL. 54, NO. 4, 557–585
SEPTEMBER 1989

LATENT VARIABLE MODELING IN HETEROGENEOUS POPULATIONS

BENGT O. MUTHÉN

GRADUATE SCHOOL OF EDUCATION, UNIVERSITY OF CALIFORNIA, LOS ANGELES

Common applications of latent variable analysis fail to recognize that data may be obtained from several populations with different sets of parameter values. This article describes the problem and gives an overview of methodology that can address heterogeneity. Artificial examples of mixtures are given, where if the mixture is not recognized, strongly distorted results occur. MIMIC structural modeling is shown to be a useful method for detecting and describing heterogeneity that cannot be handled in regular multiple-group analysis. Other useful methods instead take a random effects approach, describing heterogeneity in terms of random parameter variation across groups. These random effects models connect with emerging methodology for multilevel structural equation modeling of hierarchical data. Examples are drawn from educational achievement testing, psychopathology, and sociology of education. Estimation is carried out by the LISCOMP program.

Key words: mixtures, covariance structures, multiple-group analysis, MIMIC, LISCOMP, random parameters, multilevel, hierarchical data.

1. Introduction

In preparing this presidential address, I decided to touch on not only what I have done but also some of what I am doing and would like to see done in the future, both in terms of my own research and that of other psychometricians. Before going into the specifics of my topic, the general themes will be described.

In line with my own taste, I will concentrate on applied issues: Although "applied" is a relative term meaning different things to different people, I will present more general formulas, tables, and graphs than detailed derivations of theories and proofs. A second theme is modeling. In line with my own interests, I will focus on the specification of models rather than details of estimation. In my view, too little psychometric effort is geared towards realistic modeling which naturally should precede polishing of model parameter estimation. A final general theme is the standard statistical assumption of i.i.d., the assumption of identically and independently distributed observations. I will discuss analysis approaches that relax one or both of these assumptions. The presentation will also involve effects of ignoring i.i.d. violations, both in terms of distortions of regular analysis that maintains this assumption and, more importantly, in terms of information not uncovered by regular analysis.

Before getting into specific modeling issues, I will give a general description of my topic, including an outline of the content of the sections.

Presidential address delivered at the Psychometric Society meetings in Los Angeles, USA and Leuven, Belgium, July 1989. The research was supported by Grant No. SES-8821668 from the National Science Foundation and by Grant No. OERI-G-86-003 from the Office for Educational Research and Improvement, Department of Education. I thank Leigh Burstein, Mike Hollis, Linda Muthén, and Albert Satorra for helpful discussions and Tammy Tam, Jin-Wen Yang, Suk-Woo Kim, and Lynn Short for computational assistance. Designs were created by Arlette Collier, Rita Ling and Jennifer Edic-Bryant.

Requests for reprints should be addressed to Bengt O. Muthén, Graduate School of Education, University of California, Los Angeles, California, 90024-1521.

0033-3123/89/1200-pa89$00.75/0

2. Population Heterogeneity

In interacting with substantive researchers during recent years on issues of psychometric modeling, I have encountered an interesting common theme, namely that of population heterogeneity. Data are frequently analyzed as if they were obtained from a single population, although it is often unlikely that all individuals in our sample have the same set of parameter values.

There are many important examples of population heterogeneity in psychometric modeling with latent variables. In educational achievement modeling with factor analysis and item response theory, the homogeneity assumption is unrealistic when applied to a sample of students with varying instructional background. A good example is modeling of mathematics achievement for U.S. eighth grade students, where widely varying curricula or tracks are being followed, placing more or less emphasis on topics such as algebra and geometry (see, e.g., Muthén, 1988a, 1989a; Muthén, Kao, & Burstein, in press). In studies of attitudes and opinions the homogeneity assumption of standard measurement models may not be realistic across subsets of the group studied. For instance, in survey research the validity and reliability of certain items can be expected to vary across subgroups defined by race, gender, region, and issue salience (see, e.g., Converse, 1964; Hollis & Muthén, 1988; Schuman & Presser, 1981). In public health research such as psychiatric epidemiology, surveys frequently are concerned with data that come from a mixture of "normal" and "abnormal" subjects, for example individuals who have and have not suffered from a "major depressive disorder" (Eaton & Bohrnstedt, 1989).

An alternative view to homogeneity is that data come from a mixture of populations with their own sets of parameter values. This relates to statistical modeling called finite mixture analysis (see, e.g., Everitt & Hand, 1981). However, the situations we will consider are in one sense often simpler than those of finite mixture analysis, since the mixing population membership is assumed known with no need for estimating the mixing proportions. In another sense, our situations are often more complex than that of mixtures, since we will attempt to use population-specific variables that are auxiliary to the variables and relationships of primary interest.

With this general premise as a starting point, section 3 elaborates in more detail the consequences of analyzing a mixture of populations with regular covariance structure models for latent variables, that is in the tradition of Jöreskog (1978). Given the distortions that are shown in this section, section 4 begins by considering if regular multiple-group latent variable analysis (Jöreskog, 1971; Sörbom, 1974) is a sufficient solution. In section 4.1, the alternative of MIMIC modeling is described. Section 4.2 briefly outlines the LISCOMP analysis framework (Muthén, 1987), both since this naturally contains MIMIC modeling for continuous and categorical response variables, and since this framework will be shown to encompass the new analysis developments discussed in section 6. Section 4.3 analyzes two real-data examples with new types of MIMIC modeling for heterogeneous populations. In section 5, a transition is made from these traditional fixed-effects latent variable models to models with random parameters for hierarchical data. Section 6.1 discusses new types of models that utilize random parameter descriptions of heterogeneity, section 6.2 considers the implications for the likelihood of normally distributed response variables, and section 6.3 discusses analysis. Section 7 concludes with an outline of future possibilities.

3. Latent Variable Mixtures with Varying Levels

To be more specific about the kinds of heterogeneity to be discussed, consider the following latent variable models, also discussed in Muthén (1989b). In line with regular covariance structure modeling, assume the linear factor analysis measurement model for a set of p interval-scaled response variables y in G groups (populations), $g = 1, 2, \ldots, G$,

$$\mathbf{y}_g = \boldsymbol{\nu}_g + \boldsymbol{\Lambda}\boldsymbol{\eta}_g + \boldsymbol{\varepsilon}_g, \tag{1}$$

where $\boldsymbol{\nu}$ and $\boldsymbol{\Lambda}$ contain measurement intercept and slope (loading) parameters, $\boldsymbol{\eta}$ is an m-vector of factors, and $\boldsymbol{\varepsilon}$ is a vector of residuals. Assume that $E(\boldsymbol{\eta}_g) = \boldsymbol{\alpha}_g$, $V(\boldsymbol{\eta}_g) = \boldsymbol{\Psi}$, $V(\boldsymbol{\varepsilon}_g) = \boldsymbol{\Theta}$, and that

$$E(\mathbf{y}_g) = \boldsymbol{\nu}_g + \boldsymbol{\Lambda}\boldsymbol{\alpha}_g = \boldsymbol{\mu}_g; \tag{2}$$

$$V(\mathbf{y}_g) = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta} = \boldsymbol{\Sigma}, \tag{3}$$

so that the variable means vary across groups whereas the covariance matrix does not.

Psychometric research has covered issues of invariance in the factor model when selecting groups of individuals from an overall population for which a model such as (1) holds. For example, Meredith (1964) utilized Pearson-Lawley selection results to study cases where the measurement parameters of $\boldsymbol{\nu}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\Theta}$ are invariant under certain assumptions (see also Muthén & Jöreskog, 1983). This is a result that provides a rationale for multiple-group factor analysis with restrictions of measurement invariance (see, e.g., Jöreskog, 1971; Sörbom, 1974). The reverse situation of studying the covariance structure of an overall population where a certain factor analysis model holds in several subgroups appears to have received far less attention. Indeed, some users of structural equation models may be under the mistaken impression that when a certain simple structure holds in each of several groups, it also holds for the total sample.

Referring to the situation of a mixture of normal distributions with common $\boldsymbol{\Sigma}$ and mixture proportions w, we may generalize the two-group result of Johnson and Kotz (1972) to G groups following (1), (2), and (3):

$$\boldsymbol{\Sigma}_M = \boldsymbol{\Sigma} + \sum_{g=1}^{G} w_g(\boldsymbol{\mu}_g - \boldsymbol{\mu}_M)(\boldsymbol{\mu}_g - \boldsymbol{\mu}_M)', \tag{4}$$

where the subscript M represents parameters of the mixture distribution, $\boldsymbol{\Sigma}$ is the common covariance matrix for each group, and

$$\boldsymbol{\mu}_M = \sum_{g=1}^{G} w_g \boldsymbol{\mu}_g. \tag{5}$$

In general, the second term on the right-hand-side of (4) is such that the model that holds for $\boldsymbol{\Sigma}$ does not hold for $\boldsymbol{\Sigma}_M$. Across-group level heterogeneity may distort the structure. Hence, even when a covariance structure model is true for each group, it may not be true for the mixture.

It is interesting to also consider the special case where complete across-group invariance of measurement parameters holds. Since $\boldsymbol{\Lambda}$ and $\boldsymbol{\Theta}$ have already been taken to be group-invariant above, this means that we are now adding the assumption of invariant measurement intercept parameters ν. In the case of equal $\boldsymbol{\nu}_g$'s, (4) simplifies as

$$\boldsymbol{\Sigma}_M = \boldsymbol{\Lambda}\left(\boldsymbol{\Psi} + \sum_{g=1}^{G} w_g \boldsymbol{\alpha}_g \boldsymbol{\alpha}_g'\right)\boldsymbol{\Lambda}' + \boldsymbol{\Theta}, \tag{6}$$

where we standardize to $\sum_{g=1}^{G} w_g \boldsymbol{\alpha}_g = \mathbf{0}$. Here, we note that the expression in parentheses consists of the sum of the factor covariance matrix common to each group and a component representing variation in factor levels across groups. In terms that will be used further on, we may describe these two components as within and between factor covariance matrices. If the common factor covariance matrix is unrestricted, a regular covariance structure analysis of $\boldsymbol{\Sigma}_M$ could fit the same model structure as in each of the G groups and take the expression in parentheses as the factor covariance matrix, with $\boldsymbol{\Lambda}$ and $\boldsymbol{\Theta}$ as the remaining parameter arrays. Hence, we would consider a factor covariance matrix that is not the correct one for any of the groups. We can therefore conclude that even when the model is true for the mixture, it does not have the same parameter values as in each group. This fact is well-known in special cases such as reliability assessment. Assuming a one-factor model for (6) and defining the reliability of an observed variable as the amount of variation explained by the factor, the increase in reliability due to increased factor variance going from the group to the mixture is in line with the classic results of Lord and Novick (1968, pp. 129–131) on effects of group heterogeneity and selection on test reliability.

Two examples give interesting illustrations of the two different cases considered, with and without invariance of the measurement intercepts.

Example 1. Consider a confirmatory factor analysis model specified as a general-factor, specific-factors variance components model for nine variables. As shown in Table 1 the factor loading pattern is such that the general factor influences all variables with possibly different loadings, and the three specific factors each influence three variables and with equal loadings fixed at one. The factors are taken to be uncorrelated, the variance of the first factor is fixed at one, and the variances of the three specific factors are free parameters. This model has a variance component interpretation in that it partitions the variance in each variable into parts due to the general factor (λ^2) and the corresponding specific factor (the specific factor variance; Gustafsson, 1988, in press). The true parameter values are general factor loadings of 0.5, specific factor variances of 0.09, and residual variances of 0.66, resulting in unit observed variable variances. The resulting ratios of variance component contributions of general to total (G/T) and of specific to general (S/G) are also given in Table 1.

Assume now that this model, with the above parameter values, holds for each of two groups. Suppose that these groups are only different in terms of their factor means with no difference for the general factor, a difference of 0.3 for the first two specific factors, and a difference of 0.3 for the third specific factor. The value 0.3 corresponds to one standard deviation of each of the specific factors. Assume that we are analyzing a mixture of observations from the two groups with the mixing proportions 2/3 and 1/3. This situation can be studied by forming the population covariance matrix for the mixture by (4) and analyzing this matrix by regular maximum-likelihood structural modeling software, taking this matrix as the sample covariance matrix. The results are very interesting. The standard chi-square fit measure produced by regular software is zero with 24 degrees of freedom so that the estimated covariance matrix exactly reproduces the input matrix. In this example, measurement intercept invariance holds so that the covariance matrix for the mixture obeys (6). If the factor covariance matrix were unrestricted, the Equation (6) "between-group" addition to the factor covariance matrix clearly could be absorbed. However, here the factor covariance matrix is restricted. The addition of the between group factor covariance matrix nevertheless gets absorbed with the assistance of the general factor loading estimates. Hence, we would have no indication of misfit in this case. Despite the fact that the model holds in the mixture, the parameter values obtained are not those of the two groups. The right-most

TABLE 1

A General-Factor, Specific-Factors Variance Components Model for Nine Variables

Loading Matrix					True Values(%)		Fitted Values(%)	
	G	S_1	S_2	S_3	G/T	S/G	G/T	S/G
λ_1	1	0	0		25	36	28	31
λ_2	1	0	0		25	36	28	31
λ_3	1	0	0		25	36	28	31
λ_4	0	1	0		25	36	28	31
λ_5	0	1	0		25	36	28	31
λ_6	0	1	0		25	36	28	31
λ_7	0	0	1		25	36	19	84
λ_8	0	0	1		25	36	19	84
λ_9	0	0	1		25	36	19	84

Note: G stands for general factor, S stands for specific factor, and T stands for total.

columns of Table 1 show the resulting variance component ratios when fitting the mixture covariance matrix. A distorted picture results. Note that for the last three variables a much inflated specific factor contribution is observed.

Example 2. Consider now a one-factor model where the measurement intercepts are not invariant. Figure 1 shows a diagram of nine observed variables measuring a single factor. The variables of x_1 and x_2 are dichotomous representing two grouping variables (e.g., gender and ethnicity). Together with the product x_1x_2, these three background variables allow for differences in levels for each of the corresponding four groups. For example, the direct arrow from x_1 to y_1 allows the measurement intercept to be different for y_1 in the $x_1 = 0$ and $x_1 = 1$ categories. Similarly, the measurement intercept of y_7 varies over the four groups. Note that the diagram states that conditional on group membership, the y variables are uncorrelated when holding the factor constant, so for each group a one-factor model holds. It is also clear that in the mixture of the four groups, this is not true; y_1 and y_7 correlate over and above what the factor accounts for.

A numerical example indicates the magnitude of the possible distortion. Assume that for each of the four groups, all loadings are 0.7, the factor variance is 1.0, all residual variances are 0.51, giving unit y variances. Consider a mixture of the four groups in the proportions 3/10, 1/10, 1/10, and 5/10. Assume direct effects from each of the three background variables to y_1 and y_7 of size 1.0, the standard deviation of the y's, and assume for simplicity that there are no effects from the background variables to the

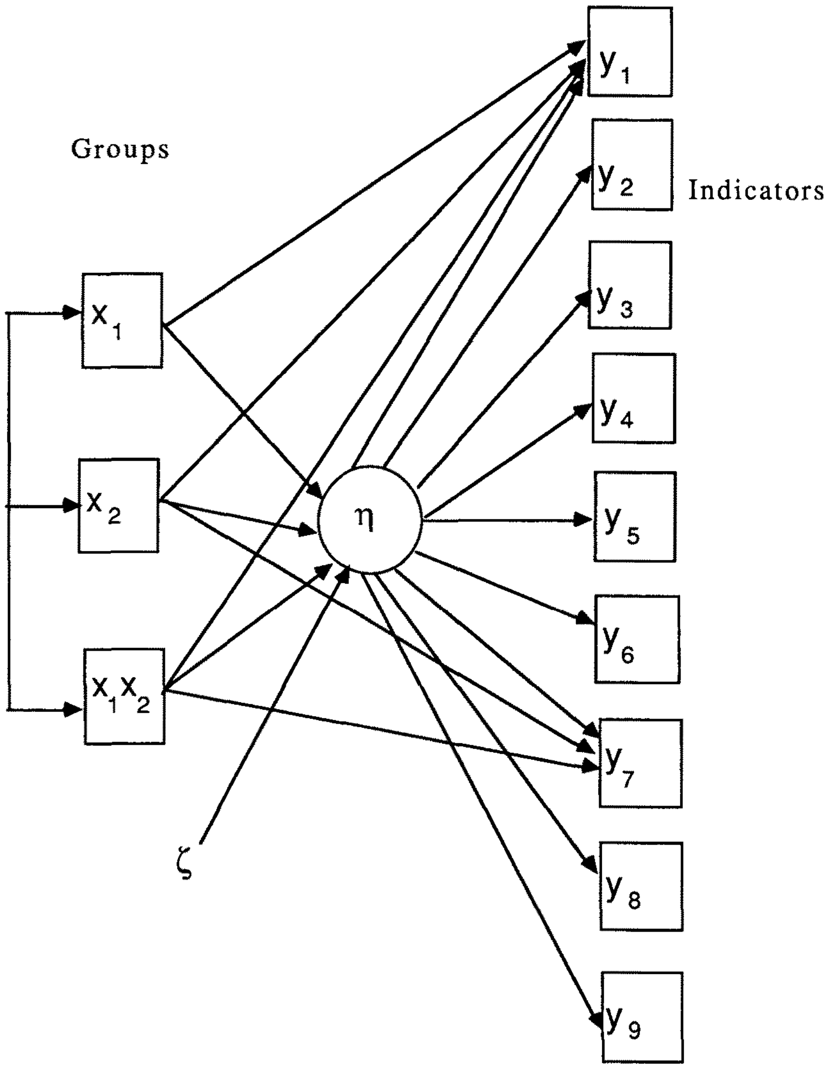

FIGURE 1.
A one-factor model with intercepts varying across groups.

factor. Fitting a one-factor model to Σ_M by maximum-likelihood estimation gives a chi-square of 14.9 with 27 degrees of freedom assuming a sample size of 500. Since a population covariance matrix has been fitted this is not a standard chi-square test value, but it gives a noncentrality parameter value from which the power of rejecting the one-factor model can be calculated using the method of Satorra and Saris (1985). The rejection power for the sample of 500 is 0.54, indicating that in more than half of the samples from this mixture, the one-factor model would be rejected. And this rejection occurs despite the fact that the one-factor model is true in each of the four groups. We also note that fitting a two-factor model to Σ_M fits perfectly, where y_1 and y_7 measure the second artificial factor arising from group-heterogeneity in measurement intercepts.

It might be thought that the group-variation in parameter values could be captured in regular latent variable analysis using multiple-group structural modeling (Jöreskog, 1971; Sörbom, 1974, 1982). As powerful as this technique is, however, it can not capture common forms of heterogeneity sufficiently well. There are two important requirements of such an analysis that are often not fulfilled. One is that sizeable samples are available for each group. We would want enough observations in each group to be able to compute stable correlation coefficients or variances and covariances. At the very least, we would like to have more observations in each group than variables. A second requirement is that group membership can be properly viewed as a fixed variable with finite levels to which inference is drawn. In what follows, examples will be presented that indicate important application areas where heterogeneity does not come in these forms.

In passing we may also note that regular multiple-group analysis also assumes that group membership does not vary across the variables. Muthén (1988b; 1989a) considers modeling in such a case with item-specific group-membership arising due to dichotomous opportunity-to-learn information for a set of mathematics test items, where the difficulty of each item is shifted by these opportunity-to-learn variables.

4. An Alternative Attempt at Capturing Heterogeneity: MIMIC Analysis

The path diagram of Figure 1 suggests an interesting alternative approach to capturing heterogeneity. The diagram has the form of a so-called MIMIC model (multiple indicators, multiple causes; see, e.g., Hauser & Goldberger, 1971). In such a model one or more latent variables intervene between observed background variables x predicting a set of observed response variables y. This section will describe in general terms how MIMIC modeling can detect and describe heterogeneity, put this modeling in a general framework, and give two examples.

4.1 MIMIC Modeling of Heterogeneity

MIMIC modeling may be seen as a way of investigating a hypothesized measurement model (a factor analysis model) for a set of response variables y capturing a set of factors. The factors and the y's are predicted by a set of regressors x that may be viewed as the covariates the y's are conditioned on. A generic MIMIC model is shown in Figure 2 for the case of a single factor.

Although conditioning on a set of x variables may appear to forfeit the latent variable modeling objective of finding invariant measurement structures, the inclusion of a set of relevant x variables provides MIMIC modeling with important extra information about such a measurement model. This enables an investigation of hypotheses of construct validity and invariance across subpopulations. First of all, predictors of the factors can be studied with respect to differential predictive strength, and at the same time give a stronger test of dimensionality. Second, of particular importance for categorical response variables, the inclusion of x variables enables a data-driven specification of the latent variable distributions. If the x's are not normal, the latent variables they predict will not be normal, and hence, the latent response variables customarily specified to underlie the y's will not be normal. In Muthén (1989c) this fact was utilized to compute so called non-normal tetrachoric correlations for the response variables. These correlations were estimated from the MIMIC model and then subjected to regular exploratory factor analysis. Third, and of immediate interest to us here, is the possibility of heterogeneity detection and modeling (also, see Muthén, 1988a, 1988b; Hollis & Muthén, 1988). Heterogeneity can be studied in two ways. In both cases, we assume that we are considering grouping variables among the x's.

First, we may allow for across-group variation in factor means. In line with the mixture cases studied in section 3, we consider

$$E(\boldsymbol{\eta}_g \mid \mathbf{x}_g) = \boldsymbol{\Gamma}\mathbf{x}_g; \tag{7}$$

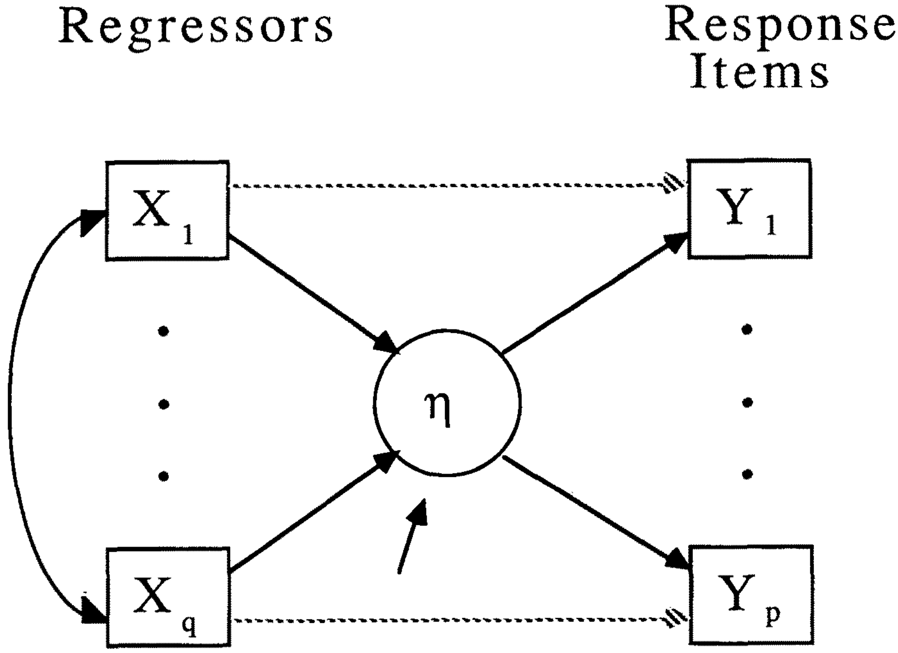

FIGURE 2.
A one-factor MIMIC model.

$$V(\boldsymbol{\eta}_g \mid \mathbf{x}_g) = \boldsymbol{\Psi}, \tag{8}$$

where the g subscript varies across the group observations, $\boldsymbol{\Gamma}$ is a matrix of regression coefficients, and $\boldsymbol{\Psi}$ is a covariance matrix. Although the covariance matrices of the factors η are assumed constant across the groups, the means are allowed to vary. For example, in Example 1, the two groups could be represented by a single dichotomous variable x that allows for the given factor mean differences. While the regular factor analysis of the covariance matrix involving only the response variables would lead to the biases shown for this example, a MIMIC analysis would essentially analyze the "within" covariance matrix, pooling across the two groups, thereby avoiding the bias. This appears to be an underutilized method (see, however, Keesling & Wiley, 1974) and one that does not appear to have been previously available for categorical data (see, however, Mislevy, 1985, 1987).

Second, the MIMIC approach allows for across-group heterogeneity in measurement intercepts. In the MIMIC context this is handled by allowing direct effects (broken arrows in Figure 2) from x's to y's. As was the case in Example 2 depicted in Figure 1, x variables corresponding to groups are thereby able to shift the level of the measurement intercepts. Specifying a standard MIMIC model with no direct effects as a base-line model, the need for including such direct effects can be detected by model modification techniques and the noninvariance accounted for where needed.

In contrast to multiple-group analysis, the MIMIC approach is restricted to modeling under the assumption of a group-invariant covariance matrix for the observed response variables, conditional on grouping variables represented by the x's. But with insufficient sample sizes for multiple-group analysis, this may be the best alternative. At least the levels of the variables are allowed to vary. Also, relative to multiple-group modeling, it is easy to accommodate a more fine-tuned categorization of the sample, using many groups.

4.2 The LISCOMP Framework

Before turning to specific examples of MIMIC heterogeneity modeling, it is instructive to consider how this fits into a general modeling framework. This framework has been utilized in the author's program "LISCOMP (Analysis of linear structural equations with a comprehensive measurement model)," described in more detail in Muthén (1983, 1984, 1987). The program was used in all examples in this article.

Let $\mathbf{y}^*$ be a p-dimensional vector of latent, continuous response variables for which a standard linear measurement model for m factors η holds,

$$\mathbf{y}^* = \boldsymbol{\nu} + \boldsymbol{\Lambda\eta} + \boldsymbol{\varepsilon}. \tag{9}$$

The y^* variables correspond to p observed y variables, which may be dichotomous (also, see Muthén, 1978, 1979; Muthén & Christoffersson, 1981), ordered polytomous (Muthén, 1983, 1984; Olsson, 1979; Olsson, Drasgow, & Dorans, 1982), continuous and censored (Muthén, 1985, in press), continuous nonnormal (see, e.g., Browne, 1982, 1984; Muthén, 1989d), and continuous normal (see, e.g., Muthén, 1987). A set of $C - 1$ threshold parameters link each latent response variable y^* to its corresponding observed y, where y has C categories. For a continuous unlimited y, we take $y = y^*$, while a censored y is only observed as y^* between censoring points.

A set of linear structural equations are also specified,

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}x + \boldsymbol{\zeta}, \tag{10}$$

where $\mathbf{x}$ represents a set of q observed variables, $\mathbf{B}$ and $\boldsymbol{\Gamma}$ are structural regression parameters, and $\boldsymbol{\zeta}$ is a vector of residuals. In the LISCOMP framework, this model is considered for G groups.

Without $\mathbf{x}$ variables ($q = 0$), the above model specification includes all regular structural equation models for continuous y's (see, e.g., Bentler, 1980, 1983; Jöreskog, 1973, 1978), using y and η to represent both independent and dependent indicators and factors. Multiple-group analysis with level structures (thresholds, means, or intercepts) for categorical and other nonnormal y's provides unique features. The inclusion of the x variables allows for the possibility of a regression-based analysis. This unique LISCOMP feature allows for the less restrictive assumption of conditional normality, given $\mathbf{x}$. Assuming conditional normality for $\mathbf{y}^*$ given $\mathbf{x}$, LISCOMP modeling considers, for each of G groups, the components of

$$E(\mathbf{y}^* \mid \mathbf{x}) = \boldsymbol{\pi}_1 + \boldsymbol{\Pi}_2\mathbf{x}; \tag{11}$$

$$V(\mathbf{y}^* \mid \mathbf{x}) = \boldsymbol{\Pi}_3, \tag{12}$$

where

$$\boldsymbol{\pi}_1 = \boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha}, \tag{13}$$

$$\boldsymbol{\Pi}_2 = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}, \tag{14}$$

$$\boldsymbol{\Pi}_3 = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})'^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}. \tag{15}$$

A corresponding set of thresholds is also considered for categorical y variables. In LISCOMP, the components of (11) and (12) are broken down into the three parts of (13) through (15), levels ($\boldsymbol{\pi}_1$), slopes ($\boldsymbol{\Pi}_2$), and correlations ($\boldsymbol{\Pi}_3$), each of which may be used alone or in combination with other parts. If there is no mean structure, or threshold structure with categorical variables, the $\boldsymbol{\pi}_1$ structure need not be included. If there are no x variables, the conditioning on $\mathbf{x}$ is vacuous and $\boldsymbol{\Pi}_2$ disappears. If only a correlation (covariance) structure is of interest, the Π_3 structure is used. For example, for a MIMIC model with categorical response, as will be exemplified in section 4.3, x's are present and the slope (Π_2) and residual correlation (Π_3) components would normally be fitted since this would consider the full implications of the model. However, we will also show an example where the analysis focuses on a residualized correlational structure, in which case only the Π_3 structure is used.

For the case of normally distributed response variables y, the estimation of MIMIC models was considered by Jöreskog and Goldberger (1975). It was pointed out that when applying maximum-likelihood, the same fitting function was obtained when assuming joint normality of $\mathbf{y}$ and $\mathbf{x}$ as when assuming normality of $\mathbf{y}$ conditional on $\mathbf{x}$ (in which case the x's need not be normally distributed; see also Jöreskog, 1973, p. 94). Muthén (1979) studied maximum-likelihood estimation for the case of dichotomous response variables and Muthén (1984) considered the generalization to ordered polytomous responses using a limited-information generalized least-squares estimator. There is also a choice in the categorical case between assuming joint or conditional normality, although here the assumption refers to the latent response variables (y^*'s) underlying each of the observed categorical y variables. Assuming joint normality leads to the use of latent variable correlations: tetrachoric, polychoric, biserial, polyserial, and tobit (Muthén, 1985, in press). A fact that has not been emphasized enough was pointed out in Muthén (1983, 1984), namely that this assumption is unnecessarily restrictive and, as opposed to the normal response variable case, does not give the same estimates as the assumption of conditional normality. The conditional normality approach leads to a structural analysis of regression coefficients, that is a regression-based as opposed to a correlation-based approach.

Muthén (1987, chap. 6) gives an overview of LISCOMP estimation in one or several groups with limited-information generalized least-squares (GLS) for categorical variables and nonnormal continuous variables and with normal theory GLS and ML for normally distributed variables. Briefly stated, the statistics to be analyzed for group g ($g = 1, 2, \ldots, G$) may be assembled in the vector $\mathbf{s}_g$ and the corresponding population entities in the vector $\boldsymbol{\sigma}_g$. The weighted least-squares estimator

$$F = \sum_{g=1}^{G} (\mathbf{s}_g - \boldsymbol{\sigma}_g)' \mathbf{W}_g^{-1} (\mathbf{s}_g - \boldsymbol{\sigma}_g), \tag{16}$$

is applied to the simultaneous analysis of G independent samples, where $\mathbf{W}_g$ may be chosen as an approximation to the asymptotic covariance matrix of $\mathbf{s}_g$ to yield a limited-information GLS estimator. The fitting function F of (16) is used for categorical and other non-normal response variables y, including censored variable estimators, multiple-group ADF (Muthén, 1985, 1989d), and regression-based analysis. When the y variables are normally distributed (and the regression-based approach is not used), the weight matrix of the GLS estimator simplifies and F reduces to

$$F^* = \sum_{g=1}^{G} \frac{(N_g - 1) \operatorname{tr} [(\boldsymbol{\Sigma}_g - \mathbf{S}_g)\mathbf{S}_g^{-1}]^2 + N_g \operatorname{tr} [\mathbf{S}_g^{-1}(\bar{\mathbf{y}}_g - \boldsymbol{\mu}_g)(\bar{\mathbf{y}}_g - \boldsymbol{\mu}_g)']}{N}, \tag{17}$$

where N_g is the sample size of group g, tr stands for trace, $\mathbf{S}_g$ is the usual sample covariance matrix, $\bar{\mathbf{y}}_g$ is the sample mean vector, and N is the total sample size. For normally distributed y variables we may also use the maximum-likelihood estimator (also, see section 6), modifying (17) as

$$F^{**} = \sum_{g=1}^{G} \frac{N_g[\log|\Sigma_g| + \operatorname{tr}(\Sigma_g^{-1}\mathbf{T}_g) - \log|\mathbf{S}_g| - p]}{N}, \tag{18}$$

where $\mathbf{T}_g = \mathbf{S}_g + (\bar{\mathbf{y}}_g - \boldsymbol{\mu}_g)(\bar{\mathbf{y}}_g - \boldsymbol{\mu}_g)'$.

4.3 MIMIC Heterogeneity Examples

Example 3. Consider the following analysis of depression and anxiety based on data from Baltimore and Durham (Eaton & Bohrnstedt, 1989). In Muthén (1989c), a factor analysis was carried out for a set of dichotomously scored symptom items administered within the Epidemiological Catchment Area Program to approximately 3,500 individuals at each of the two sites. Three clearly interpretable factors, termed Phobic Anxiety, Somatic Anxiety, and Depression, were found for a subset of 27 items. It is of interest to take this analysis further to study differences in factor and symptom levels across groups (also, see Muthén, 1989b). It is well known for example that these symptoms vary in prevalence across gender and ethnicity. Different sites may also show differences due to their varying sociodemographic composition. Indeed, judging from the mixture examples of section 3, the heterogeneity of levels across groups makes the factor analysis results questionable.

A regular structural modeling approach would in this case carry out an analysis of each gender × ethnicity × site group separately and then test for measurement invariance in a multiple-group analysis where factor mean differences could be estimated. In this case, however, the large total sample size is not large enough to support a stable estimation of the correlations for all of the groups. This is because the symptoms are rare with few people admitting to both of any pair of symptoms. This therefore gives an example of what was discussed at the end of section 3, namely that regular latent variable modeling is often inadequate due to lack of sufficient sample sizes for the groups.

MIMIC modeling provides a solution in this case. The groups may be represented by a set of dummy x variables, in this case seven variables representing the cells of the $2 \times 2 \times 2$ table formed by gender (female or not) × ethnicity (black, nonblack) × site (Durham or not). Since the response variables are dichotomous, the regression-based approach assuming conditional normality for the y^*'s given the x's is advantageous. This approach also allows for variation in the factor means across the eight groups and can also capture non-invariance in the measurement intercepts for those items where this is warranted.

The MIMIC model estimation was carried out by LISCOMP's limited-information generalized-least squares estimator for dichotomous response from (16). A base-line model with no direct effects was first applied, followed by a model modification where the need for direct effects where detected with the help of modification indices of first-order derivatives for fixed parameters (see Muthén, 1989b). The estimated model gave a factor pattern that actually closely agreed with what was found in factor analysis of the tetrachoric correlations when analyzing the y's only, so here the possibility of distortion was not realized. Other LISCOMP estimates are given in Table 2, where the columns correspond to the dummy x variables, the first three rows correspond to the three factors, and the last 27 rows correspond to the items. Entries are unstandardized structural regression coefficients. Here, we will only be concerned with sign and significance of the effects.

From the first three rows we note, for example, that Females have a higher Somatic Anxiety factor level than Males (0.53) and that this is the only factor that has a

TABLE 2

Estimated Effects of Groups on Factors and Items

	FEMALE	DURHAM	BLACK	FB	FD	BD	FBD
ANXIETY	.53*	-.18	.51*	.15	-.09	-.09	-.11
SOMATIC	.22*	.29*	-.06	-.21	-.07	-.06	.41*
DEPRESS	.24*	-.07	-.06	.30*	.03-	.05	-.19
ANIMALS	.00	.00	.00	.00	.00	.00	.00
BREATH	.00	.00	.00	.00	.00	.00	.00
BUGS	.08	.22	.00	.05	-.06	.00	.00
CLOSED	.00	.00	-.02*	.00	.00	.00	.00
CROWD	.00	.00	.00	.00	.00	.00	.00
CRYING	.51*	.00	.00	.00	.00	.00	.00
DIZZY	.00	.00	.00	.00	.00	.00	.00
PANIC	.00	.00	.20*	.00	.00	.00	.00
APPETITE	.00	.00	.00	.00	.00	.00	.00
SLEEP	.00	.00	.00	-.15*	.00	.00	.00
SLOW	-.17*	.00	.00	.00	.00	.00	.00
INTEREST	.00	.00	.00	.00	.00	.00	.00
TIRED	.00	.00	-.12*	.00	.00	.00	.00
WORTHLESS	.00	.00	.00	.00	.00	.00	.00
THINKING	.00	.00	.00	.00	.00	.00	.00
DEATH	-.12*	.00	.00	.00	.00	.00	.00
HEIGHTS	-.32*	-.08	-.30*	.32	-.06	-.12	-.02
HOPELESS	.00	.00	.00	.00	.00	.00	.00
NERVOUS	.15*	.00	-.24*	.00	.00	.00	.00
GOINGOUT	.00	.00	.00	.00	.00	.00	.00
HEARBEAT	.00	.00	.00	.00	.00	.00	.00
PUBTRANS	.00	.00	.00	.00	.00	.00	.00
DYSPH	.00	.00	.00	.00	.00	.00	.00
STORMS	.25*	.00	.00	.00	.00	.00	.00
TUNBRI	.00	.00	.00	.00	.00	.00	.00
WATER	.00	.00	.00	.00	.00	.00	.00
WEAK	.00	.00	.00	.00	.00	.00	.00

*Significant on 1% level

site-specific effect. The remaining rows show interesting instances of direct effects from the group variables to the items. Consider, for example, the item Crying. Crying is an indicator of depression. Females have a higher depression factor mean and are therefore expected to admit to this symptom to a higher degree than Males. The positive direct effect (0.51), however, shows that the Crying level for Females is elevated beyond what would be predicted by the factor increase. This means that the Crying indicator does not show measurement invariance across gender. Consider next (Fear of) Heights. This item is an indicator of Anxiety. Blacks (and Females) have a higher Anxiety factor mean that non-Blacks (Males). However, the negative direct effect shows that the expected corresponding increase in Fear of Heights prevalence is not fully realized for Blacks (Females); the Fear of Heights item has different measurement characteristics in these groups.

In this situation there appears to be no good alternative to the regression-based MIMIC approach. What is sometimes tried is to do a regular factor analysis of the items for the total sample, compute factor scores, and calculate means for the different groups. As pointed out in Muthén (1989b), this analysis ignores the different forms of level heterogeneity and does not properly fit even the univariate distributions of the items. It also suffers from the usual estimation errors of factor scores. Indeed, for this example, results emerge that conflict with those presented above.

Example 4. As a somewhat different example of MIMIC modeling of heterogeneity, consider a general-factor, specific-factors model of the kind discussed in Example 1 of section 3 (also, see Gustafsson, 1988; in press). In this case we analyze 40 dichotomous mathematics achievement items from the U.S. eight grade sample of the Second International Mathematics Study (Crosswhite, Dossey, Swafford, McKnight, & Cooney, 1985). In previous analyses by Muthén, Burstein, Gustafsson, Webb, Kim, and Short (1989) several uncorrelated specific factors with narrow item domains corresponding to instructional segments had been identified orthogonal to a general math achievement factor influencing all items. As in Example 1, there was an interest in studying the variance contribution of the specific factors relative to the general factor. These analyses were based on confirmatory factor analysis using tetrachoric correlations. However, the analyses did not take into account the strong degree of heterogeneity in instructional background and eighth grade curricula, which realistically would cause strong variation in factor levels. There were about 200 classes in the sample of 3,724 students. These classes had been categorized into Remedial, Typical, Enriched, and Algebra. To at least account for group differences across these categories, three dummy variables were used as x variables in a regression-based MIMIC analysis. The dummy variable gender was also added to reflect possible gender heterogeneity.

The analysis in this case is carried out in two steps. First, a multivariate probit regression is carried out to get estimated slopes and an estimated correlation matrix. Here, the factor model is not imposed, but $\mathbf{\Pi}_2$ and $\mathbf{\Pi}_3$ of (14) and (15) are unrestricted. Second, the corresponding estimated correlations among the y^* variables are computed (also, see Muthén, 1989c). These correlations are then subjected to a confirmatory analysis with the model used for the regular tetrachorics. In effect, then, the MIMIC-type analysis works with a pooled-within tetrachoric correlation matrix.

The results are given in Table 3. The rows correspond to the specific factors and the entries are percentage variation in the items (y^* variables), where the general factor entry corresponds to the average variance contribution for the items of that specific factor. The two left-most columns of percentages refer to the regular tetrachoric analysis for the y's only, while the two right-most columns refer to the MIMIC-based analysis.

The regular analysis of y's shows for example that the variance contribution of the specific factor Angular (for angular measurement items in geometry) is about 40 percent (11/28) of that of the general factor. In the MIMIC-based analysis, the corresponding contribution is about 60% (13/22). Relative to the regular analysis, the MIMIC-based analysis consistently decreases the general factor variance contribution, while the specific factor contributions are about the same or slightly higher. The decrease in the general factor contribution is natural since conditioning on class type, the MIMIC-based analysis to some extent controls for selection effects that are presumably most strongly related to the general factor; part of the variation in the general factor which

TABLE 3

Item Variance Component Estimates (%) for a General Factor, Specific-Factor Variance Component Model For 8th-Grade Math

Specific Factors	Analysis of Correlations for			
	y only		y given x	
	General	Specific	General	Specific
VISUAL	37	4	32	4
ALGTRANS	32	9	25	10
DECIMAL	28	7	22	7
PERCENT	35	10	29	10
ESTIMATE	32	9	25	10
APPROX	23	7	18	7
NUMCOMP	35	4	27	5
QUADRLA	39	9	32	10
ANGULAR	28	11	22	13

is observed in the regular analysis is due to between-group differences (compare with (4) in section 3).

This example shows the benefits of an analysis of a pooled-within matrix, in this case a type of tetrachoric correlation matrix (correlations estimated for the y^* variables assumed to underlie the dichotomous y's). Still, only a very crude representation of instructional differences is obtained by the class type dummy variables. Further reductions in heterogeneity can realistically be expected from using more refined groupings. For example, it would be interesting to analyze this variance component model while controlling for level differences across classrooms. With 200 classrooms this obviously leads to a very cumbersome analysis with a proliferation of parameters even in a MIMIC framework. Furthermore, classrooms have been randomly selected and a random effects representation of classroom differences may be more appropriate than the MIMIC approach of a fixed (non-random) parameter for each factor mean in each classroom. This is what we described as the second type of inadequacy of regular latent variable analysis at the end of Section 3. Leaving the MIMIC framework, we will now turn our attention to the modeling of heterogeneity using random parameter techniques applied to latent variable structural models.

5. Random Parameter Modeling

Regular structural equation modeling with latent variables, either with groups represented in a multiple-group analysis or a MIMIC analysis with groups corresponding to x variable combinations, take a fixed effects approach. Parameters are viewed as varying over a finite number of groups. An alternative is the random effects approach, where parameters are viewed as continuous, random variables. The random parameter approach is well-established in random coefficient regression in, for example, econometrics and agriculture, including of course variance component estimation; see, for example Swamy (1970), Maddala (1977), and Mundlak (1978). Recently, these tech-

TABLE 4

Overview of the Multilevel Research Field

Regressor	Regression parameters: Fixed	Random
Fixed	Ordinary regression	Random coefficient regression
Random	Simultaneous equations	Scarcity of research
Random and latent	Latent variable structural models	Scarcity of research

Source: Muthén & Satorra (1989). Multilevel Aspects of Varying Parameters in Structural Models. In R. D. Bock (Ed), Multilevel analysis of educational data.

niques have become more accessible and more popular in educational research through the development of so called multilevel regression models and software for the analysis of hierarchical data; see, for example, Aitkin and Longford (1986), Burstein (1980), Burstein, Kim, Delandshere (1988), de Leeuw and Kreft (1986), Goldstein (1986, 1987), Longford (1987), Mason, Wong, and Entwistle (1984), and Raudenbush and Bryk (1988). Here it is recognized that much educational data is not obtained as a simple random sample but in a hierarchical fashion with students sampled within schools and classrooms.

In a chapter of a recent book on multilevel analysis (Bock, 1989a), Muthén and Satorra (1989) attempt to structure the multilevel research field from the point of view of structural models with latent variables. Table 4, taken from their chapter, gives a simple 3 × 2 classification of relevant modeling approaches.

For all entries, we may consider the essence of the modeling as regressions, where the measurement part of a structural model represents regressions on latent regressors. The columns correspond to regression parameters that are treated as fixed versus random. Let us consider each row in turn. First, consider regressors that are fixed, or random but conditioned upon. With fixed parameters, this is the standard regression case, with the generalization to random parameters described in the references just mentioned. Random regressors occur when a variable takes the role of both an independent and dependent variable in a simultaneous equation system or path analysis model. Although techniques are well-developed for fixed parameters, there is a scarcity of research in the random parameter case. For the case of random and latent regressors, such as in factor analysis, there is also a scarcity of research in the random parameter case. The remainder of this article will consider the situation of this bottom right cell. Although some work exists in these areas, such as de Leeuw (1985), Schmidt and Wisenbaker (1986), Goldstein and McDonald (1988), McDonald and Goldstein (1988), and Muthén and Satorra (1989), much more remains to be done.

Muthén and Satorra (1989) identify two distinguishing features of multilevel modeling:

1. We are considering a heterogeneous population. Individuals are observed within different groups, and it is realistic to assume that individuals of different groups obey different response processes and relationships between variables.
2. We do not have independence among all our observations. It is realistic to assume that individuals within a group share certain influencing factors and hence have correlated observations.

In our discussion of heterogeneity we have so far considered the first of these two aspects, the across-group parameter variation. However, as was shown in the mathematics achievement example, Example 4 of the last section, the heterogeneity often comes in the form of hierarchical data for which the second aspect is important as well. Hence, we will now consider relaxing both of the two parts of the i.i.d. assumption as mentioned in the introduction.

6. Methods for Hierarchical Data

6.1 Latent Variable Models with Random Parameter Variation

Consider now some extensions of a model proposed in Muthén and Satorra (1989). The first model variation will be termed the Muthen-Satorra varying factor means model, which may be viewed as a parsimonious baseline model for heterogeneous groups. The Muthen-Satorra model was motivated by applications such as the mathematics achievement analysis of Example 4, where strong heterogeneity could be expected for the levels of the factors. For individual i within group g, consider the factor analysis model

$$\mathbf{y}_{gi} = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_{gi} + \boldsymbol{\varepsilon}_{gi}, \tag{19}$$

where $E(\boldsymbol{\varepsilon}_{gi}) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}_{gi}) = \boldsymbol{\Theta}$ for all g's and i's, and

$$\boldsymbol{\eta}_{gi} = \boldsymbol{\alpha}_g + \boldsymbol{\omega}_{gi}; \tag{20}$$

$$\boldsymbol{\alpha}_g = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{z}_g + \boldsymbol{\delta}_{\alpha_g}, \tag{21}$$

where $\boldsymbol{\alpha}_g$ is a group-level random component, $\boldsymbol{\omega}_{gi}$ is an individual-level random component, and $\mathbf{z}_g$ is a vector of observed group-level variables.

Conditional on group g,

$$E(\boldsymbol{\eta}_{gi} \mid g) = \boldsymbol{\alpha}_g, \tag{22}$$

$$V(\boldsymbol{\eta}_{gi} \mid g) = V(\boldsymbol{\omega}). \tag{23}$$

In line with the mixture and MIMIC assumptions of sections 3 and 4, the factor mean is allowed to vary across groups and the factor covariance matrix is not. Note that $\boldsymbol{\alpha}_g$ is a random variable vector in this specification. Considering the special case of no group-level variables z and a single factor, only the single parameter $V(\delta_\alpha)$ would be needed to capture the group heterogeneity in factor levels. In contrast, the regular fixed effects latent variable approach would use $G - 1$ parameters (one parameter is fixed for identification purposes).

This modeling specifies correlated factor scores for individuals i and j in group g,

$$\text{Cov}\,(\boldsymbol{\eta}_{gi}, \boldsymbol{\eta}_{gj'}) = \boldsymbol{\Gamma} V(\mathbf{z})\boldsymbol{\Gamma}' + V(\boldsymbol{\delta}_{\alpha}). \tag{24}$$

We obtain

$$\mathbf{y}_{gi} = \boldsymbol{\nu} + \boldsymbol{\Lambda}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{z}_g + \boldsymbol{\delta}_{\alpha_g}) + \boldsymbol{\Lambda}\boldsymbol{\omega}_{gi} + \boldsymbol{\varepsilon}_{gi}, \tag{25}$$

and

$$V(\mathbf{y}) = \boldsymbol{\Sigma}_W + \boldsymbol{\Sigma}_B, \tag{26}$$

with

$$\boldsymbol{\Sigma}_W = \boldsymbol{\Lambda} V(\boldsymbol{\omega})\boldsymbol{\Lambda}' + \boldsymbol{\Theta}; \tag{27}$$

$$\boldsymbol{\Sigma}_B = \boldsymbol{\Lambda}\boldsymbol{\Gamma} V(\mathbf{z})\boldsymbol{\Gamma}'\boldsymbol{\Lambda}' + \boldsymbol{\Lambda} V(\boldsymbol{\delta}_{\alpha})\boldsymbol{\Lambda}'. \tag{28}$$

Consider again the special case of no group-level variables z. We note that a regular analysis of $V(\mathbf{y})$ would consider the estimation of $\boldsymbol{\Lambda}$, $V(\boldsymbol{\omega}) + V(\boldsymbol{\delta}_{\alpha})$, and $\boldsymbol{\Theta}$ so that the within and between components of the factor covariance matrix would be confounded (compare (4)). These components will become separable given within- and between covariance information as described below.

In line with our discussion in previous sections, the varying factor means model of (19) through (28) may also be augmented by allowing for across-group variation in the measurement intercepts so that as for $\boldsymbol{\alpha}_g$ in (21), $\boldsymbol{\nu}_g$ is a random vector written as a function of group-level variation $\mathbf{z}_g$ and $\boldsymbol{\delta}_{\nu_g}$. If there are no z's, this adds a matrix component $V(\boldsymbol{\delta}_{\nu})$ to $\boldsymbol{\Sigma}_B$ in (28). Assuming that this matrix is diagonal and there are no z's, the $\boldsymbol{\Sigma}_W$ and $\boldsymbol{\Sigma}_B$ structures are the same with equal factor loading matrices.

Considering a factor analysis model structure for both $\boldsymbol{\Sigma}_W$ and $\boldsymbol{\Sigma}_B$ suggests interpretating the model as simultaneously fitting factor models to the within and between covariance matrix parts, with certain parameters possibly being invariant across levels. This interpretation focuses on analysis on multiple levels and their interactions, commonly referred to as multilevel modeling (see Burstein, 1980; McDonald & Goldstein, 1988; Schmidt & Wisenbaker, 1986), while our model description has focussed on improving the usual individual-level modeling by allowing for across-group parameter heterogeneity in variable levels.

The Muthén-Satorra assumption of equality of loading matrices in $\boldsymbol{\Sigma}_W$ and $\boldsymbol{\Sigma}_B$ is clearly seen in the two terms involving $\boldsymbol{\Lambda}$ in (25). Although this equality constraint may be a good approximation in many factor analysis contexts, certain applications may require different effects of group and individual-level factor components on the observed variables. In this context, it is interesting to note some early "multilevel factor analysis modeling" attempts. Cronbach (1976) reanalyzed Bond-Dykstra data on readiness measures across classrooms using within and between covariance matrices. Using a similar approach, Härnqvist (1978) analyzed primary mental ability scores for students in Grades 4 through 9 across classes and districts. In both instances, there were indications of different factor structures at the different levels.

It is also clear that the above specification is directly generalizable to structural relations among the factors, where there may be group-level variation in structural equation intercepts, factor means, and measurement intercepts. The Muthen-Satorra specification of using different parameter matrices $V(\boldsymbol{\omega})$ and $V(\boldsymbol{\delta})$ for within and between factor variation then generalizes to allowing for different structural slopes and structural residual variances on the within and between levels. Muthén (1989e) considers further modeling, estimation, and computation aspects.

In all these model variations, this article focuses on estimation of the structural parameters underlying $\boldsymbol{\mu}_z$, $\boldsymbol{\mu}_y$, $\boldsymbol{\Sigma}_{zz}$, $\boldsymbol{\Sigma}_{yz}$, $\boldsymbol{\Sigma}_W$, and $\boldsymbol{\Sigma}_B$. In terms of multilevel modeling we note that such parameters are in Bayesian terms referred to as hyperpopulation parameters (see e.g., Lindley & Smith, 1972). In line with the Bayesian approach, we may also be interested in estimating each group's factor value $\boldsymbol{\alpha}_g$, assuming "exchangeability" of the G groups. This is a generalization of factor score estimation to the group level. In the regression model context, such estimation is often carried out by Empirical Bayes (see e.g., Braun, Jones, Rubin, & Thayer, 1983; Rubin, 1980, 1983).

6.2 The Likelihood for Hierarchical Data

To understand the differences between regular latent variable analysis of observations from heterogeneous groups and multilevel modeling with randomly varying parameters it is instructive to consider the likelihood of the data. Assume $g = 1, 2, \ldots, G$ independently observed groups with $i = 1, 2, \ldots, N_g$ individual observations within group g. Let $N = \Sigma N_g$ be the total number of observations. As before, let $\mathbf{z}$ and $\mathbf{y}$ represent group- and individual level variables, respectively. In line with Tiao and Tan (1965) and others, arrange the data vector for which independent observations are obtained as

$$\mathbf{d}_g' = (\mathbf{z}_g', \mathbf{y}_{g1}', \mathbf{y}_{g2}', \ldots, \mathbf{y}_{gN_g}'), \tag{29}$$

where we note that the length of $\mathbf{d}_g$ varies across groups. The mean vector and covariance matrix of $\mathbf{d}_g$ are

$$\boldsymbol{\mu}_{d_g}' = [\boldsymbol{\mu}_z, \mathbf{1}_{N_g}' \otimes \boldsymbol{\mu}_y']; \tag{30}$$

$$\boldsymbol{\Sigma}_{d_g} = \begin{bmatrix} \boldsymbol{\Sigma}_{zz} & \text{symmetric} \\ \mathbf{1}_{N_g} \otimes \boldsymbol{\Sigma}_{yz} & \mathbf{I}_{N_g} \otimes \boldsymbol{\Sigma}_W + \mathbf{1}_{N_g}\mathbf{1}_{N_g}' \otimes \boldsymbol{\Sigma}_B \end{bmatrix}, \tag{31}$$

where $\boldsymbol{\mu}_z$ and $\boldsymbol{\mu}_y$ are the mean vectors of $\mathbf{z}$ and $\mathbf{y}$, $\otimes$ denotes the Kronecker product, $\mathbf{1}_{N_g}$ denotes a vector of N_g unit elements, $\mathbf{I}_{N_g}$ is the identity matrix of dimension N_g, $\boldsymbol{\Sigma}_{yz}$ contains the covariances between $\mathbf{z}$ and $\mathbf{y}$, and $\boldsymbol{\Sigma}_W$ and $\boldsymbol{\Sigma}_B$ are covariance matrices for $\mathbf{y}$ (compare section 6.1). Assuming multivariate normality of $\mathbf{d}_g$, the maximum-likelihood (ML) estimator minimizes the function

$$F^{**} = \sum_{g=1}^{G} \{\log |\boldsymbol{\Sigma}_g| + (\mathbf{d}_g - \boldsymbol{\mu}_g)'\boldsymbol{\Sigma}_g^{-1}(\mathbf{d}_g - \boldsymbol{\mu}_g)\}. \tag{32}$$

Consider the case of no group-level variables z. From (31) we note that if $\boldsymbol{\Sigma}_B = \mathbf{0}$, the $\mathbf{y}_{gi}$ observations are independent not only across g but also across i. In this case, F^{**} reduces to the regular structural modeling ML fitting function for N identically and independently distributed observations on $\mathbf{y}$. The matrix $\boldsymbol{\Sigma}_B$ allows for explicit modeling of the correlations among observations within homogeneous groups (see Point 2 in the Muthén-Satorra quote of section 5). If $\boldsymbol{\Sigma}_B \neq \mathbf{0}$, we do not have N independent observations.

The general likelihood expression of (32) was studied by McDonald and Goldstein (1988) in the context of multilevel structural equation modeling, with the somewhat different aim of providing multilevel latent variable path analysis with relations both within and across levels. The work of McDonald and Goldstein is an important contribution in that they worked out simplifications to the computationally unwieldy expression of (32). It turns out that the g-th term reduces to simpler matrix expressions, the size of which do not involve the number of observations within groups but only the number of variables,

$$\begin{aligned}
&\log|\boldsymbol{\Sigma}_{zz}| + (N_g - 1)\log|\boldsymbol{\Sigma}_W| + \log|\boldsymbol{\Sigma}_g| \\
&+ \operatorname{tr}\{[\boldsymbol{\Sigma}_{zz}^{-1} + N_g\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{yz}'\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Sigma}_{yz}\boldsymbol{\Sigma}_{zz}^{-1}](\mathbf{z}_g - \boldsymbol{\mu}_z)(\mathbf{z}_g - \boldsymbol{\mu}_z)'\} \\
&- 2N_g \operatorname{tr}\{\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{yz}'\boldsymbol{\Sigma}_g^{-1}(\bar{\mathbf{y}}_g - \boldsymbol{\mu}_y)(\mathbf{z}_g - \boldsymbol{\mu}_z)'\} \\
&+ \operatorname{tr}\left\{\boldsymbol{\Sigma}_W^{-1}\sum_{i=1}^{N_g}(\mathbf{y}_{gi} - \boldsymbol{\mu}_y)(\mathbf{y}_{gi} - \boldsymbol{\mu}_y)'\right\} \\
&- N_g \operatorname{tr}\{[\boldsymbol{\Sigma}_W^{-1} - \boldsymbol{\Sigma}_g^{-1}](\bar{\mathbf{y}}_g - \boldsymbol{\mu}_y)(\bar{\mathbf{y}}_g - \boldsymbol{\mu}_y)'\}, \qquad (33)
\end{aligned}$$

where

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}_W + N_g\boldsymbol{\Sigma}_{B.z}, \qquad (34)$$

$$\boldsymbol{\Sigma}_{B.z} = \boldsymbol{\Sigma}_B - \boldsymbol{\Sigma}_{yz}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{yz}'. \qquad (35)$$

It is interesting to consider the special case of balanced data, that is, equal N_g's across groups. Also, assume no group-level z variables and $\boldsymbol{\mu}_y$ unrestricted. In this case, the fitting function F^{**} of (32) simplifies considerably and reduces to

$$\begin{aligned}
F^* = G\log|n^{-1}\boldsymbol{\Sigma}_W + \boldsymbol{\Sigma}_B| + (N - G)\log|\boldsymbol{\Sigma}_W| + G\operatorname{tr}[(n^{-1}\boldsymbol{\Sigma}_W + \boldsymbol{\Sigma}_B)^{-1}\mathbf{S}_B] \\
+ (N - G)\operatorname{tr}[\boldsymbol{\Sigma}_W^{-1}\mathbf{S}_{PW}], \qquad (36)
\end{aligned}$$

where n is the common group size, $\mathbf{S}_B$ is a between-group sample covariance matrix

$$\mathbf{S}_B = G^{-1}\sum_{g=1}^{G}(\bar{\mathbf{y}}_g - \bar{\mathbf{y}})(\bar{\mathbf{y}}_g - \bar{\mathbf{y}})', \qquad (37)$$

and $\mathbf{S}_{PW}$ is the regular pooled-within sample covariance matrix,

$$\mathbf{S}_{PW} = (N - G)^{-1}\sum_{g=1}^{G}\sum_{i=1}^{n}(\mathbf{y}_{gi} - \bar{\mathbf{y}}_g)(\mathbf{y}_{gi} - \bar{\mathbf{y}}_g)'. \qquad (38)$$

Although most psychometricians have apparently been unaware of this, the possibility of fitting multilevel covariance structure models with the ML fitting function of (36) has in fact been available for 20 years! In his unpublished dissertation, Schmidt (1969) studied multivariate random effects models and provided a general computer program for maximum likelihood estimation using a Fletcher-Powell optimization algorithm. Schmidt and Wisenbaker (1986) studied the slightly more general case of structural equation modeling and presented an example which we will reanalyze shortly.

Consider now a simple reformulation of the general likelihood expression in (32) through (35), due to Muthén (1989e). Note that the group-level variation in y may be expressed as

$$V(\bar{\mathbf{y}}_g) = N_g^{-1}\boldsymbol{\Sigma}_W + \boldsymbol{\Sigma}_B, \tag{39}$$

and consider the between-group covariance matrix

$$\boldsymbol{\Sigma}_{gg} = \begin{bmatrix} \boldsymbol{\Sigma}_{zz} & \text{symmetric} \\ \boldsymbol{\Sigma}_{yz} & N_g^{-1}\boldsymbol{\Sigma}_W + \boldsymbol{\Sigma}_B \end{bmatrix}. \tag{40}$$

The fitting function F^{**} may then be written as

$$F = \sum_{d=1}^{D} G_d\{\log|\boldsymbol{\Sigma}_{dd}| + \mathrm{tr}[\boldsymbol{\Sigma}_{dd}^{-1}\mathbf{S}_{Bd}]\}$$

$$+ (N - G)\{\log|\boldsymbol{\Sigma}_W| + \mathrm{tr}[\boldsymbol{\Sigma}_W^{-1}\mathbf{S}_{PW}]\}, \tag{41}$$

where D is the number of distinct group sizes, G_d is the number of groups of a particular size, and

$$\mathbf{S}_{Bd} = G_g^{-1} \sum_{k=1}^{G_d} (\mathbf{v}_k - \boldsymbol{\mu})(\mathbf{v}_k - \boldsymbol{\mu})', \tag{42}$$

for $\mathbf{v}_k' = (\mathbf{z}_k, \bar{\mathbf{y}}_k')$, $\boldsymbol{\mu}' = (\boldsymbol{\mu}_z', \boldsymbol{\mu}_y')$.

This way of writing the likelihood has interesting connotations. Consider first the balanced case. Here, we have a single distinct group size ($D = 1$) and one $\mathbf{S}_B$ matrix. In the case of $\boldsymbol{\mu}$ unrestricted, F simplifies to that $\mathbf{S}_B$ need not be centered around the population mean as in (42), but can instead be centered at the overall sample mean. When there are no group-level variables z, the resulting expression is equivalent to Schmidt's Equation (36). The division of the F expression into two lines suggests optimization via existing multiple-group structural equation modeling software for ML estimation, such as LISCOMP! F can in fact be optimized in a two-group simultaneous analysis of a between group with G observations and sample covariance matrix $\mathbf{S}_B$ and a within group with $N - G$ observations and a sample covariance matrix $\mathbf{S}_{PW}$ (compare with (18)). As described by Muthén (1989e), any of the model variations discussed in connection with the Muthén-Satorra model can be handled in this framework.

In the unbalanced case with $\boldsymbol{\mu}$ unrestricted, there are several group sizes ($D > 1$) and several $\mathbf{S}_{Bd}$ matrices. Centering the $\mathbf{S}_{Bd}$'s at the sample mean no longer gives ML estimation, since the ML estimate of $\boldsymbol{\mu}$ is not the overall sample mean. However, given large samples, the overall sample mean may not be far from the ML estimate of $\boldsymbol{\mu}$. Using the sample mean instead of $\boldsymbol{\mu}$ yields a simple and perhaps quite reasonable estimator that can still be handled in a multiple-group fashion with only slightly modified structural modeling software. These matters are studied further in Muthén (1989e).

6.3 Analysis of Hierarchical Data

The new modeling possibilities presented by multilevel models create a need to consider an appropriate model testing sequence. To clearly understand the data structure, a step-wise sequence of increasingly more complex models is recommended as follows.

Step 1. To check that the hypothesized structural model is at all reasonable, it is useful to first analyze the regular sample covariance matrix by regular models. Although failing to take multilevel aspects into account may create a certain amount of distortion in this analysis, the usual kinds of misspecifications are presumably of larger magnitude and are the first ones that should be cleared up. If $\Sigma_B = \mathbf{0}$ in the multilevel setting, this analysis gives the same ML estimates as a multilevel structural analysis as pointed out in connection with the likelihood for hierarchical data. If the model fit is "in the ballpark", go to Step 2.

Step 2. The second step is to fit the Σ_W structure to the pooled-within matrix $\mathbf{S}_{PW}$ by regular structural modeling. This provides a check of the appropriateness of the model and takes level heterogeneity into account, much like the MIMIC analyses discussed in section 4. For the balanced case, this gives the same ML estimates as a multilevel structural model with Σ_B free (unrestricted). If the model fit is in the ball park, go to Step 3.

Step 3. Test significance of between-group variation (and within-group correlatedness) by multilevel analysis comparing fit of $\Sigma_B = \mathbf{0}$ versus Σ_B free in either of two ways:

(i) using Σ_W free;
(ii) using the structure of Σ_W applied in Steps 1 and 2.

If Σ_B is significantly different from zero, go to Step 4.

Step 4. Use multilevel modeling applying the model of Steps 1 and 2 to Σ_W and use as a base-line model for Σ_B the Muthén-Satorra varying factor means model.

Step 5. If needed, relax the Step 4 restrictions on Σ_B as appropriate, for example in terms of the Muthén-Satorra loading matrix invariance and allowing intercept variation. Or, if Step 4 gives a well-fitting model, test further restrictions related to Σ_B, in steps ending with $\Sigma_B = \Sigma_W$.

An example will now be provided for a balanced case using LISCOMP.

Example 5. Figure 3 shows a path diagram for a structural model studied by Schmidt and Wisenbaker (1986). It refers to data from the National Longitudinal Study, with observations on a national sample of high school students graduating in 1972. In this case, student observations are obtained hierarchically within high schools. Schmidt and Wisenbaker studied balanced data with 13 students per school. It is of interest to study heterogeneity in the variable levels using the Muthén-Satorra varying factor level specification. We may also investigate heterogeneity in measurement intercepts. Related modeling was studied in Schmidt and Wisenbaker. As was the case in their analyses, the between level variation in Foreign Language Courses caused nonconvergence and this variable was deleted in our analyses, so that the factor of Verbal Skill Courses is identically equal to the observed variable English Courses. The between and pooled-within sample covariance matrices used in the analyses are given in Table 5, using the definitions of (37) and (38). Here Sex is coded as 1 for Males and Ethnicity is coded as 1 for Whites.

Table 6 presents the results of applying our model testing sequence to the NLS model of Figure 3. From Step 3 it is clear that there is a strong need for a multilevel model that does not restrict Σ_B to zero. The simple Muthén-Satorra type multilevel model fits the data sufficiently well and there appears to be no need to include further parameters in the model, such as intercept variation. The factor analysis model of Muthén and Satorra is of course here reparameterized as a structural model so that different within and between group structural equations are considered.

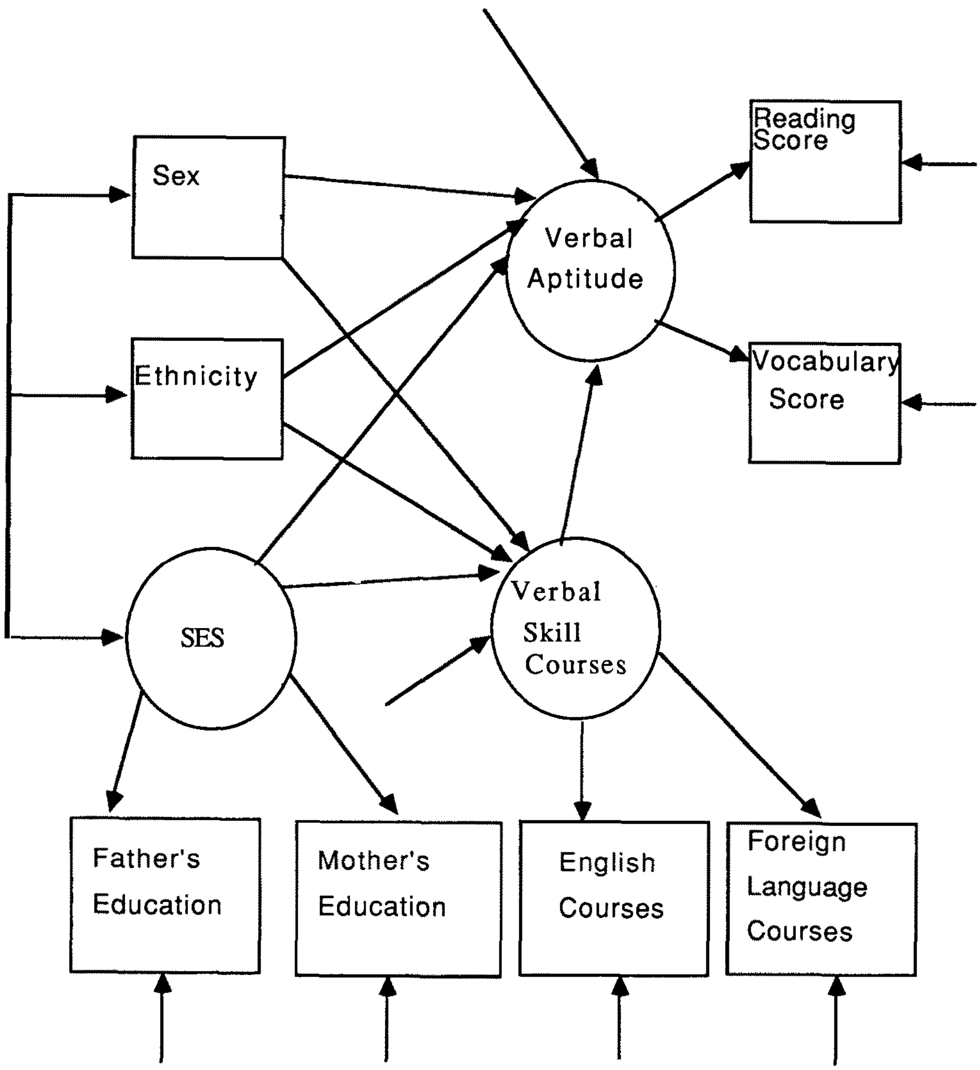

FIGURE 3.
Model For National Longitudinal Study Data (Source: Schmidt & Wisenbaker, 1986).

Table 7 gives the estimated model, conveniently presented in terms of individual (within)- and group (between)-level parameters, in line with Schmidt and Wisenbaker (1986). The structural variances under the heading "Between" reflect parameter heterogeneity in terms of across-school variation in factor (or variable) means. We note that the between-school structural variance of Verbal Aptitude is about 18% of the total Verbal Aptitude variance. It is sometimes argued that the between component is relatively small due to the fact that the within component partly contains errors of measurement. In this case, however, the variance ratio pertains to the variance of the error-free factor. For the endogeneous SES factor the between-school contribution is somewhat larger, about 26% of the total SES variance. The structuring of the within- and between factor covariance matrices in terms of structural equations gives an interesting interpretation of the across-school heterogeneity. The differences in R^2 for the within and between regressions of Verbal Aptitude are striking, 20% versus almost 80%, perhaps in line with the increase in "ecological correlations" as compared to

TABLE 5

Between and Within Sample Covariance Matrices for NLS Data

Between

Foreign	5.9277							
English	0.0860	6.1225						
Reading	2.2541	-0.4485	5.0019					
Vocab.	2.2231	-0.5654	3.5118	3.7543				
Sex	-.0574	.0163	-.0032	.0056	.0338			
Ethnicity	.0160	-.0109	.3239	.2672	.0042	.0699		
FAED	1.2078	-.1890	1.4349	1.3478	.0082	.0938	1.2466	
MOED	.7726	-.0752	.9743	.8658	.0106	.0618	.7242	.6471

Within

Foreign	16.7090							
English	0.8619	4.5062						
Reading	7.0760	1.1563	24.1637					
Vocab.	5.8330	1.0543	12.6414	15.8329				
Sex	-.1989	-.0037	-.0054	-.0347	-.2589			
Ethnicity	.0895	.0282	.3662	.2708	.0033	.1073		
FAED	1.6774	.3029	2.2122	1.9478	.0163	.0888	4.2629	
MOED	1.2817	.2524	1.6933	1.5199	.0206	.0585	1.7908	2.8427

Source: Schmidt & Wisenbaker (1986). Foreign and English have been divided by 40.

regular ones (Robinson, 1950). There are also differences in terms of value and significance of the structural coefficients. This is for example true for the influence of Verbal Courses (English courses) on Verbal Aptitude, where in contrast to the student-level relation, the Verbal Aptitude variation across schools appears not to be significantly influenced by Verbal Courses. Further tests can be made of similarities of within and between level structural regressions, but were not carried out here.

It is interesting to note that the Step 1 regular structural analysis of the regular sample covariance matrix gives results that are rather similar to those of the "Within" column, both in terms of estimates and standard errors of estimates. Referring back to the themes outlined in the introduction, a regular individual-level analysis does not give a strongly distorted picture but is, in this case, roughly correct as far as it goes. The point is that the regular analysis is incapable of uncovering interesting aspects of the data related to between-group variation.

TABLE 6

NLS Model: Tests of Fit (N = 1,300, G = 100)

	Model	χ^2	d.f.
1.	Regular analysis	3.2	7
2.	Regular on pooled-within	3.0	7
3.	Multilevel, $\Sigma_B = 0$, Σ_W free	1,363.0	28
4.	Multilevel, Muthén-Satorra	20.2	20

7. Conclusions

In terms of latent variable methodology, this article has covered a rather wide variety of techniques, although they all share the capability of uncovering various forms of population heterogeneity. It was pointed out that latent variable modeling by multiple-group, MIMIC, and multilevel analysis are useful techniques for modeling heterogeneity. For the type of applications that we have considered, MIMIC analysis was shown to be superior to multiple-group analysis. Multilevel analysis provides further flexibility and interesting possibilities will also be opened up when using multilevel analysis in combination with MIMIC and multiple-group approaches.

It seems clear that random parameter multilevel modeling techniques have further untapped potential for latent variable analysis in heterogeneous populations. This is true whether one is interested primarily in the random parameter part of the statement or the multilevel part. An immediate example is longitudinal modeling, in which case the individual takes the role of independently observed groups within which correlated observations (corresponding to students within classrooms) are obtained over time. Such modeling can recognize across-individual parameter variation and may begin to respond to the Rogosa (1987) critique of structural equation modeling being insensitive to individual differences in growth (also, see Rogosa & Willett, 1985). Some interesting initial work in this area is described in, for example, Bock (1983; 1989b), Gibbons and Bock (1987), Hedeker, Gibbons, and Waternaux (1988), and Raudenbush and Bryk (1988). See also Muthén (1989f).

Other areas are largely unexplored. For example, can this approach be used for demographic cross-classifications in surveys? Here, one can study a group defined as a certain multiway cross-classification and allow across-group differences in parameters despite the fact that very few individuals belong to some of the groups. This modeling would then recognize that individuals of the same cross-classification share certain common experiences and may have correlated observations. This also relates to properly accounting for complex sampling procedures in estimating structural models (also, see Goldstein & McDonald, 1988). Finally, the MIMIC examples of section 4 both referred to categorical response variables, while section 6 only discussed normal var-

TABLE 7

NLS Model 4 Estimates

Structural Variances

	Within		Between	
Verbal Aptitude	14.38		3.12	
Verbal Courses	4.51		5.78	
SEX	0.26	(24.5)	0.01	(2.86)
Ethnicity	0.11	(24.5)	0.06	(6.23)
SES	2.54	(10.9)	0.88	(5.18)

Structural Slopes

	Within		Between	
Verbal Aptitude on:	$R^2 = 0.20$		$R^2 = 0.79$	
	Raw*	Stand.	Raw	Stand.
Verbal Courses	0.19 (3.44)	0.11	-0.07 (-1.07)	-0.09
SEX	-0.19 (-0.83)	-0.03	-1.39 (-0.70)	-0.09
Ethnicity	2.42 (6.50)	0.21	3.14 (4.53)	0.44
SES	0.81 (7.86)	0.34	1.17 (5.81)	0.62
Verbal Courses on:	$R^2 = 0.01$		$R^2 = 0.01$	
	Raw	Stand.	Raw	Stand.
SEX	-0.03 (-0.21)	-0.01	0.86 (0.26)	0.04
Ethnicity	0.16 (0.86)	0.03	0.05 (0.04)	0.01
SES	0.12 (2.57)	0.09	-0.23 (-0.68)	-0.09

*Raw means unstandardized coefficients; Stand. means standardized.
Z values in parentheses.

iables. Much remains to be done in the area of heterogeneity analysis by multilevel models for categorical and other nonnormal data. Some initial work has been done by Wong and Mason (1985), Longford (1988), and Bock (1989b).

To conclude, one might ask if the modeling techniques discussed above, even in the outlined extensions, are sufficient in terms of capturing heterogeneity in real data. In my opinion, the answer to that question is no. Real-world applications would appear to require much more elaborate models. As an example related to the anxiety and depression example (Example 3), the diagnosis of a major depressive episode is made based on what could be measured as a set of dichotomous symptom items, where the individual has to have been sad for two weeks and have at least four of a set of eight other symptoms (Eaton & Bohrnstedt, 1989). If the sad item is not switched on, the other symptoms indicate a syndrome of a different kind. This measurement and classification situation does not correspond to a standard factor analysis model, even when making the usual allowance for the dichotomous nature of the responses. Heterogeneity is at hand, where the items obey one model with the sad item switched on and another model when it is not. This is reminiscent of the econometric switching regressions situation with different regimes (see, e.g., Maddala, 1983, pp. 283–287), although in a latent variable context. One may, for example, entertain the possibility of classifying the individual into depressed and not depressed categories when the sad item is switched on, and provide a continuous factor score when it is not. This relates to models that mix latent class and latent trait specifications, discussed by Yamamoto (1988). But even so, individual heterogeneity is likely to require more complex models.

In summary, it is safe to say that real-world applications require further development of more tailored modeling that carefully takes into account the special features of a certain subject-matter application area. This can probably only be done well in close collaboration with substantive researchers. There is a challenge, however, for methodological researchers to provide such tailored modeling within more generally applicable methods that avoid a proliferation of dataset-specific techniques.

References

Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society, Series A 149*, 1–43.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology, 31*, 19–456.

Bentler, P. M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika, 48*, 493–518.

Bock, R. D. (1983). The discrete Bayesian. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement* (pp. 103–115). Hillsdale, NJ: Erlbaum.

Bock, R. D. (1989a). *Multilevel analysis of educational data*. San Diego, CA: Academic Press.

Bock, R. D. (1989b). Measurement of human variation: A two-stage model. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 319–340). San Diego, CA: Academic Press.

Braun, H., Jones, D., Rubin, D., & Thayer, D. (1983). Empirical Bayes estimation of coefficients in the general linear model from data of deficient rank. *Psychometrika, 48*, 171–181.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.) *Topics in applied multivariate analysis*. Cambridge, MA: Cambridge University Press.

Browne, M. W. (1984). Asymptotically distribution free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62–83.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education, 8*, 158–233.

Burstein, L., Kim, K. S., & Delandshere, G. (1988). Multilevel investigation of systematically varying slopes: Issues, alternatives, and consequences. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 235–276). San Diego, CA: Academic Press.

Converse, P. E. (1964). The nature of belief systems in mass publics. In D. Apter (Ed.), *Ideology and discontent*. London, England: The Free Press.

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., Dossey, J. A., Swafford, J. O., McKnight, C. C., & Cooney, T. J. (1985). *Second international mathematics study: Summary report for the United States*. Champaign, IL: Stipes.

de Leeuw, J. (1985). *Path models with random coefficients*. Leiden, The Netherlands: University of Leiden, Department of Data Theory.

de Leeuw, J., & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics, 11*, 57–85.

Eaton, W., & Bohrnstedt, G. (1989). Introduction. Latent Variable Models for Dichotomous Outcomes: Analysis of data from the Epidemiological Catchment Area Program. *Sociological Methods & Research, 18*, 4–18.

Everitt, B., & Hand, D. J. (1981). *Finite Mixture Distributions*. New York: Chapman and Hall.

Gibbons, R., & Bock, R. (1987). Trend in correlated proportions. *Psychometrika, 52*, 113–124.

Goldstein, H. I. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika, 73*, 43–56.

Goldstein, H. I. (1987). *Multilevel Models in Educational and Social Research*. London: Oxford University Press.

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

Gustafsson, J. E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence, Vol. 4* (pp. 35–71). Hillsdale, NJ: Lawrence Erlbaum.

Gustafsson, J. E. (in press). Broad and narrow abilities in research on learning and instruction. In R. Kanfer, P. L. Ackerman, & R. Cudeck (Eds.), *Abilities, motivation, and methodology: The Minnesota symposium on learning and individual differences*. Hillsdale, NJ: Lawrence Erlbaum.

Härnquist, K. (1978). Primary mental abilities of collective and individual levels. *Journal of Educational Psychology, 70*, 706–716.

Hauser, R. M., & Goldberger, A. S. (1971). The treatment of unobservable variables in path analysis. In H. L. Costner (Ed.), *Sociological Methodology 1971* (pp. 81–177). San Francisco: Jossey-Bass.

Hedeker, D., Gibbons, R. D., & Waternaux, C. (1988, June). *Random regression models for longitudinal psychiatric data*. Paper presented at the Annual Meeting of the Psychometric Society, Los Angeles.

Hollis, M., & Muthén B. (1988). *Incorporating "Item-specific" information into covariance structure models of political attitudes and policy preferences*. Paper prepared for delivery at the 1988 Annual Meeting of the American Political Science Association.

Johnson, N., & Kotz, S. (1972). *Distribution in statistics: Continuous multivariate distributions*. New York: John Wiley & Sons.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409–426.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85–112) New York: Seminar Press.

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika, 43*, 443–477.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent vaviable. *Journal of the American Statistical Association, 70*, 351.

Keesling, J. W., & Wiley, D. E. (1974, March). *Regression models of hierarchical data*. Paper presented at the Annual Meeting of the Psychometric Society, Palo Alto, CA.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B, 34*, 1–41.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested effects. *Biometrika, 74*, 817–827.

Longford, N. T. (1988). *A quasilikelihood adaption for variance component analysis*. Princeton, NJ: Educational Testing Service.

Long, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Maddala, G. S. (1977). *Econometrics*. New York: McGraw-Hill.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge, MA: Cambridge University Press.

Mason, W. M., Wong, G. Y., & Entwistle, B. (1984). Contextual analysis through the multi-level linear model. *Sociological Methodology*, San Francisco, CA: Jossey-Bass.

McDonald, R. P., & Goldstein, H. (1988, June). *Balanced versus unbalanced designs for linear structural relations in two-level data*. Paper presented at the Annual meeting of the Psychometric Society, Los Angeles.

Meredith, W. (1964). Notes on factorial invariance. *Psychometrika, 29*, 177–185.

Mislevy, R. J. (1975). Estimation of latent group effects. *Journal of the American Statistical Association, 80*, 993–997.

Mislevy, R. J. (1987). Exploiting auxiliary information about examinees in the estimation of item parameters. *Applied Psychological Measurement, 11*, 81–91.

Mundlak, Y. (1978). Models with variable coefficients: Integration and extension. *Annales de'l INSEE, 30–31*, 483–509.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika, 43*, 551–560.

Muthén, B. (1979). A structural probit model with latent variables. *Journal of the American Statistical Association, 74*, 807–811.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics, 22*, 43–65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*, 115–132.

Muthén, B. (1985, July). *Tobit factor analysis*. Paper presented at the Fourth European Meeting of the Psychometric Society, Cambridge, England. (Forthcoming in *The British Journal of Mathematical and Statistical Psychology*).

Muthén, B. (1987). *LISCOMP: Analysis of linear structural equations with a comprehensive measurement model* [User's Guide]. Mooresville, IN: Scientific Software.

Muthén, B. (1988a). Some uses of structural equation modeling in validity studies: Extending IRT to external variables. In H. Wainer & H. Braun (Eds.), *Test Validity* (pp. 213–238). Hillsdale, NJ: Lawrence Erlbaum.

Muthén, B. (1988b). *Instructionally sensitive psychometrics: Applications to the Second International Mathematics Study*. Unpublished manuscript, University of California, Graduate School of Education, Los Angeles.

Muthén, B. (1989a). Using item-specific instructional information in achievement modeling. *Psychometrika, 54*, 385–396.

Muthén, B. (1989b). *Covariance structure modeling in heterogeneous populations: Mean-adjusted analysis*. In preparation, University of California, Graduate School of Education, Los Angeles.

Muthén, B. (1989c). Dichotomous factor analysis of symptom data. In Eaton and Bohrnstedt (Eds.), Latent Variable Models for Dichotomous Outcomes: Analysis of Data from the Epidemiological Catchment Area Program. *Sociological Methods & Research, 18*, 19–65.

Muthén, B. (1989d). Multiple-group factor analysis with non-normal continuous variables. *British Journal of Mathematical and Statistical Psychology, 42*, 55–62.

Muthén, B. (1989e). *Covariance structure analysis of hierarchical data*. Paper presented at the American Statistical Association meeting in Washington D.C.

Muthén, B. (1989f, October). *Analysis of longitudinal data using latent variable models with varying parameters*. Invited paper presented at the University of Southern California Conference on Best Methods for the Analysis of Change, Los Angeles. (Forthcoming as a chapter in a book edited by J. Horn and L. Collins)

Muthén, B. (in press). Moments of the censored and truncated bivariate normal distribution. *British Journal of Mathematical and Statistical Psychology*.

Muthén, B., Burstein, L., Gustafsson, J-E., Webb, N., Kim, S-W., Short, L. (1989). *General and specific factors in mathematics achievement data*. In preparation, University of California, Graduate School of Education, Los Angeles.

Muthén, B., & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika, 46*, 485–500.

Muthén, B., & Jöreskog •. (1983). Selectivity problems in quasi-experimental studies. *Evaluation Review, 7*, 139–173.

Muthén, B., Kao, Chih-fen, & Burstein, L. (in press). Instructional sensitivity in mathematics achievement test items: Applications of a new IRT-based detection technique. *Journal of Educational Measurement*.

Muthén, B., & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. In Bock (Ed.), *Multilevel Analysis of Educational Data* (pp. 87–99). San Diego: Academic Press. (Invited paper for the conference on *Multilevel analysis of educational data*, Princeton, NJ, April 1987)

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44*, 443–460.

Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. *Psychometrika, 37*, 337–347.

Raudenbush, S., & Bryk, A. (1988). Methodological advances in studying effects of schools and classrooms on student learning. *Review of Research in Education*, 1988.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351–357.

Rogosa, D. R. (1987). Causal models do not support scientific conclusions: A comment in support of Freedman. *Journal of Educational Statistics, 12*, 185–195.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*, 203–228.

Rubin, D. B. (1980). Using empirical Bayes techniques in the Law School Validity Studies. *Journal of the American Statistical Association, 75*, 801–827.

Rubin, D. B. (1983). Some applications of Bayesian statistics to educational data. *The Statistician, 32*, 55–68.

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50*, 83–90.

Schmidt, W. H. (1969). *Covariance structure analysis of the multivariate random effects model*. Unpublished doctoral dissertation, University of Chicago.

Schmidt, W., & Wisenbaker, J. (1986). *Hierarchical data analysis: An approach based on structural equations* (CEPSE, No. 4., Research Series). University of Michigan, Department of Counseling Educational Psychology and Special Education.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229–239.

Sörbom, D. (1982). Structural equation models with structured means. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (pp. 183–195). Amsterdam: North-Holland.

Swamy, P. A. V. B. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, 311-323.

Tiao, G. C., & Tan, W. Y. (1965). Bayesian analysis of random-effect models in the analysis of variance. I. Posterior distribution of variance components. *Biometrika, 52*, 37–53.

Wong, G. Y., & Mason, W. M. (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association, 80*, 513–524.

Yamamoto, K. (1988). *A model that combines IRT and latent class models*. Paper presented at the 1988 American Educational Research Association Annual Meeting.