Multilevel structural equation modeling

Sophia Rabe-Hesketh
Graduate School of Education & Graduate Group in Biostatistics,
University of California, Berkeley
& Institute of Education, University of London

Anders Skrondal
Department of Statistics & Methodology Institute, London School of Economics &
Division of Epidemiology, Norwegian Institute of Public Health, Oslo

Xiaohui Zheng
Graduate School of Education, University of California, Berkeley

Abstract

In conventional structural equation models, all latent variables and indicators vary between units (typically subjects) and are assumed to be independent across units. The latter assumption is violated in multilevel settings where units are nested in clusters, leading to within-cluster dependence. Different approaches to extending structural equation models for such multilevel settings are examined. The most common approach is to formulate separate within-cluster and between-cluster models. An advantage of this set-up is that it allows software for conventional structural equation models to be 'tricked' into estimating the model. However, the standard implementation of this approach does not permit cross-level paths from latent or observed variables at a higher level to latent or observed variables at a lower level, and does not allow for indicators varying at higher levels. A multilevel regression (or path) model formulation is therefore suggested in which some of the response variables and some of the explanatory variables at the different levels are latent and measured by multiple indicators. The *Generalized Linear Latent and Mixed Modeling* (GLLAMM) framework allows such models to be specified by simply letting the usual model for the structural part of a structural equation model include latent and observed variables varying at different levels.

Models of this kind are applied to the U.S. sample of the Program for International Student Assessment (PISA) 2000 to investigate the relationship between the school-level latent variable 'teacher excellence' and the student-level latent variable 'reading ability', each measured by multiple ordinal indicators.

Keywords: Multilevel structural equation models, generalized linear mixed models, latent variables, random effects, hierarchical models, item response theory, factor models, adaptive quadrature, empirical Bayes, GLLAMM

To appear in:
S. Y. Lee (Ed.). Handbook on Structural Structural Equation Models. Amsterdam: Elsevier.

1 Introduction

The popularity of multilevel modeling and structural equation modeling (SEM) is a striking feature of quantitative research in the medical, behavioral and social sciences. Although developed separately and for different purposes, SEM and multilevel modeling have important communalities since both approaches include latent variables or random effects to induce, and therefore explain, correlations among responses.

Multilevel regression models are used when the data structure is hierarchical with elementary units at level 1 nested in clusters at level 2, which in turn may be nested in (super)clusters at level 3, and so on. The latent variables, or *random effects*, are interpreted as unobserved heterogeneity at the different levels which induce dependence among all lower-level units belonging to a higher-level unit. Random intercepts represent heterogeneity between clusters in the overall response and random coefficients represent heterogeneity in the relationship between the response and explanatory variables.

Structural equation models are used when the variables of interest cannot be measured perfectly. Instead, there are either sets of items reflecting a hypothetical construct (e.g. depression) or fallible measurements of a variable (e.g. calory intake) using different instruments. The latent variables, or *factors*, are interpreted as constructs, traits or 'true' variables, underlying the measured items and inducing dependence among them. The measurement model is sometimes of interest in its own right, but relations among the factors or between factors and observed variables (the structural part of the model) are often the focus of investigation.

Importantly, *multilevel structural equation modeling*, a synthesis of multilevel and structural equation modeling, is required for valid statistical inference when the units of observation form a hierarchy of nested clusters and some variables of interest are measured by a set of items or fallible instruments. Multilevel structural equation modeling also enables researchers to investigate exciting research questions which could not otherwise be validly addressed. For instance, in this chapter we will consider an important question in education: does student ability (a student-level latent trait) depend on teacher excellence (a school-level latent trait)?

Multilevel structural equation models could be specified using either multilevel regression models or structural equation models as the vantage point. An advantage of using the multilevel regression approach taken here is that the data need not be balanced and missing data are easily accommodated.

2 Response Types

2.1 Continuous responses

Structural equation models were originally developed for continuous responses. In this case the 'response model' or 'measurement model' for subject j, relating the observed response vector $\mathbf{y}_j$ of manifest variables or indicators to the latent variables $\boldsymbol{\eta}_j$, the observed covariates $\mathbf{x}_j$, and the error terms $\boldsymbol{\epsilon}_j$ (usually representing 'unique factors'), has the general form

$$\mathbf{y}_j \;=\; \boldsymbol{\nu}_j + \boldsymbol{\epsilon}_j, \qquad \boldsymbol{\epsilon}_j \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Theta}).$$

Here $\boldsymbol{\nu}_j$ are functions of $\boldsymbol{\eta}_j$ and $\mathbf{x}_j$ (see Section 3) and $\boldsymbol{\Theta}$ is the covariance matrix of $\boldsymbol{\epsilon}_j$, usually specified as diagonal.

2.2 Non-continuous responses

2.2.1 Latent response formulation

When the responses are dichotomous or ordinal, the same model as above can be specified for latent continuous responses $\mathbf{y}_j^*$ underlying the observed responses $\mathbf{y}_j$. A threshold model links the observed response for the ith indicator to the corresponding latent response,

$$y_{ij} = s \;\text{ if }\; \kappa_{is} < y_{ij}^* \leq \kappa_{i,s+1}, \qquad s = 0, \ldots, S-1, \quad \kappa_{i0} = -\infty, \; \kappa_{iS} = \infty.$$

The threshold parameters κ_{is} (apart from κ_{i0} and κ_{iS}) can all be estimated if the mean and variance of $\mathbf{y}_j^*$ are fixed. Alternatively, two thresholds can be fixed (typically $\kappa_{i1} = 0$ and $\kappa_{i2} = 1$) for each response variable to identify the means and variances of $\mathbf{y}_j^*$.

Grouped or interval censored continuous responses can be modeled in the same way by constraining the threshold parameters to the limits of the censoring intervals. By allowing unit-specific right-censoring, this approach can be used for discrete time durations.

An advantage of the latent response formulation is that conventional models can be specified for the underlying continuous responses. By changing the distribution of $\boldsymbol{\epsilon}_j$, the latent response formulation can also be used to specify logit models. Models for comparative responses such as rankings or pairwise comparisons can be formulated in terms of latent responses conceptualized as utilities or utility differences (e.g., Skrondal and Rabe-Hesketh, 2003).

2.2.2 Generalized linear model formulation

Unfortunately, the latent response formulation cannot be used to specify Poisson models for counts. Instead, a generalized linear model formulation is typically used where the conditional expectation of the response y_{ij} for indicator i given $\mathbf{x}_j$ and $\boldsymbol{\eta}_j$ is 'linked' to the linear predictor ν_{ij} via a link function $g(\cdot)$,

$$g(\mathrm{E}[y_{ij}|\mathbf{x}_j, \boldsymbol{\eta}_j]) \;=\; \nu_{ij}. \tag{1}$$

The linear model given above for continuous responses uses an identity link whereas the latent response model for dichotomous responses can be expressed as a generalized linear model with a probit or logit link. Other possible links are the log, reciprocal and complementary log-log.

The final component in the generalized linear model formulation is the conditional distribution of the response variable given the latent and explanatory variables. The conditional distribution is a member of the exponential family of distributions; a normal distribution is typically used for continuous responses, a Bernoulli distribution for dichotomous responses and a Poisson distribution for counts. In structural equation models with several latent variables, the measurement models for different latent variables may require different links and/or distributions.

For ordinal responses, the generalized linear model formulation is modified so that the link function is applied to cumulative probabilities instead of expectations,

$$g(\mathrm{P}[y_{ij} > s|\mathbf{x}_j, \boldsymbol{\eta}_j]) \;=\; \nu_{ij} - \kappa_{i,s+1}.$$

The threshold parameters $\kappa_{i,s+1}$ could alternatively be viewed as part of category-specific linear predictors ν_{ij}^s (treating multinomial responses as multivariate), but this will not be done here.

In structural equation modeling with categorical (dichotomous or ordinal) manifest variables, the latent response formulation is predominant. In contrast, item response models are invariably specified via the generalized linear model formulation (e.g., Mellenbergh, 1994). Although Takane and de Leeuw (1987) and Bartholomew (1987) pointed out the equivalence of the two formulations for many models, the literatures are still quite separate.

In the remainder of this chapter, we will use the generalized linear model formulation because it handles more response types. In most cases we are primarily interested in the form of the linear predictors ν_{ij} and view the choice of link functions and distributions as of secondary interest. For response types that can be modeled via a latent response formulation, the model for the latent responses can be written as $\nu_{ij} + \epsilon_{ij}$.

3 Multilevel Measurement Models

3.1 Single-level factor models

Conventional single-level factor models can be specified as

$$\boldsymbol{\nu}_j \;=\; \boldsymbol{\beta} + \boldsymbol{\Lambda}\boldsymbol{\eta}_j, \qquad \boldsymbol{\eta}_j \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Psi}).$$

For observed or latent continuous responses it follows that

$$\mathbf{y}_j^* \;=\; \boldsymbol{\beta} + \boldsymbol{\Lambda}\boldsymbol{\eta}_j + \boldsymbol{\epsilon}_j, \qquad \boldsymbol{\epsilon}_j \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Theta}). \tag{2}$$

Here $\boldsymbol{\nu}_j$ and $\mathbf{y}_j^*$ are I-dimensional vectors with elements corresponding to the indicators, $\boldsymbol{\beta}$ is a vector of intercepts, $\boldsymbol{\Lambda}$ a matrix of factor loadings, $\boldsymbol{\eta}_j$ a m-dimensional vector of common factors and $\boldsymbol{\epsilon}_j$ a vector of unique factors. The covariance structure of the latent responses becomes

$$\boldsymbol{\Sigma} \equiv \mathrm{Cov}(\mathbf{y}_j^*) \;=\; \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}, \tag{3}$$

which is called a 'factor structure'. The factor model can be specified either directly as in (2) or via the above covariance structure.

An example of an 'independent clusters' two-factor model (where each indicator measures one and only one common factor) for $I\!=\!6$ is

$$\underbrace{\begin{bmatrix} \nu_{1j} \\ \nu_{2j} \\ \nu_{3j} \\ \nu_{4j} \\ \nu_{5j} \\ \nu_{6j} \end{bmatrix}}_{\boldsymbol{\nu}_j} = \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\begin{bmatrix} \eta_{1j} \\ \eta_{2j} \end{bmatrix}}_{\boldsymbol{\eta}_j},$$

where the first factor is measured by the first three indicators and the second factor by the remaining indicators.

A path diagram for this model is given in Figure 1 where circles represent latent variables and rectangles observed variables. For continuous observed responses the long arrows represent linear relations between the responses and the common factors and the short arrows represent linear relations between the responses and the unique factors. For other response types the long arrows represent possibly nonlinear relations depending on the link function and the short arrows represent residual variability, following for instance a Bernoulli or Poisson distribution.

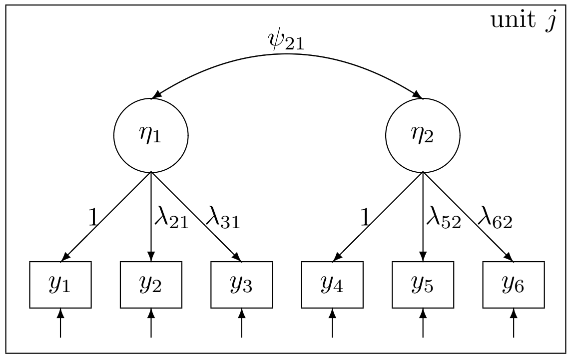

Figure 1: Independent clusters two-factor model

Factor models have a similar structure to random coefficient models as has been pointed out in the context of growth curve models (Skrondal, 1996), item response models (Rijmen *et al.*, 2003; De Boeck and Wilson, 2004) and more generally in Skrondal and Rabe-Hesketh (2004). A two-level random coefficient model can be written as

$$\boldsymbol{\nu}_j \;=\; \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{\eta}_j,$$

where the design matrix of known constants $\mathbf{Z}_j$ (varying over clusters j) takes the place of the parameter matrix of unknown factor loadings $\boldsymbol{\Lambda}$ (constant over clusters j).

Generalized linear latent and mixed models (GLLAMMs) (Rabe-Hesketh *et al.*, 2004) unify factor models and random coefficient models by allowing each latent variable to multiply a term of the form $\mathbf{Z}_j\boldsymbol{\lambda}$, where $\mathbf{Z}_j$ is a design matrix and $\boldsymbol{\lambda}$ a parameter vector. The GLLAMM formulation of the independent clusters two-factor model discussed previously is as follows:

$$\underbrace{\begin{bmatrix}\nu_{1j}\\ \nu_{2j}\\ \nu_{3j}\\ \nu_{4j}\\ \nu_{5j}\\ \nu_{6j}\end{bmatrix}}_{\boldsymbol{\nu}_j} = \underbrace{\begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\\ \beta_4\\ \beta_5\\ \beta_6\end{bmatrix}}_{\boldsymbol{\beta}} + \eta_{1j}\underbrace{\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&0\\0&0&0\\0&0&0\end{bmatrix}}_{\mathbf{Z}_{1j}}\underbrace{\begin{bmatrix}1\\ \lambda_{21}\\ \lambda_{31}\end{bmatrix}}_{\boldsymbol{\lambda}_1} + \eta_{2j}\underbrace{\begin{bmatrix}0&0&0\\0&0&0\\0&0&0\\1&0&0\\0&1&0\\0&0&1\end{bmatrix}}_{\mathbf{Z}_{2j}}\underbrace{\begin{bmatrix}1\\ \lambda_{52}\\ \lambda_{62}\end{bmatrix}}_{\boldsymbol{\lambda}_2},$$

where $\mathbf{Z}_{1j}$ and $\mathbf{Z}_{2j}$ are design matrices containing only fixed constants whereas $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are vectors of factor loadings. If the factor loadings are known constants, the products $\mathbf{Z}_{1j}\boldsymbol{\lambda}_1$ and $\mathbf{Z}_{2j}\boldsymbol{\lambda}_2$ become column vectors, giving a random coefficient model.

When viewing factor models as similar to random coefficient models, it is useful to describe the indicators as level-1 units and the subjects as level-2 units (or clusters). In the remainder of this chapter, we will therefore denote higher levels in which subjects are nested as level-3, level-4, etc.

3.2 Two-level factor models

Multilevel factor models are typically required if the subjects of interest are clustered in some way, for instance students clustered in schools.

3.2.1 Within and between formulation

The two-level factor model for subjects j in clusters k is often formulated in terms of the within-cluster and between-cluster covariance matrices, $\boldsymbol{\Sigma}_W$ and $\boldsymbol{\Sigma}_B$, respectively (e.g., Longford & Muthén, 1992; Poon & Lee, 1992; Longford, 1993; Linda, Lee & Poon 1993).

For continuous observed or latent responses, the following two-stage formulation can be used

$$\begin{aligned} \mathbf{y}^*_{jk} &\sim \mathrm{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_W) \\ \boldsymbol{\mu}_k &\sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_B). \end{aligned} \quad (4)$$

Here, $\boldsymbol{\mu}$ is the overall intercept and $\boldsymbol{\mu}_k$ are cluster-specific intercepts. Factor structures of the form in (3) are then specified for the two covariance matrices

$$\boldsymbol{\Sigma}_W = \boldsymbol{\Lambda}^{(2)}\boldsymbol{\Psi}^{(2)}\boldsymbol{\Lambda}^{(2)\prime} + \boldsymbol{\Theta}^{(2)},$$

and

$$\boldsymbol{\Sigma}_B = \boldsymbol{\Lambda}^{(3)}\boldsymbol{\Psi}^{(3)}\boldsymbol{\Lambda}^{(3)\prime} + \boldsymbol{\Theta}^{(3)}.$$

Here we have used the superscript (2) to denote subject-level variables and parameters and (3) to denote the cluster-level counterparts. For consistency with the literature, we call the model a two-level factor model although we think of items as level-1 units, subjects as level-2 units and clusters as level-3 units.

The two-level factor model can alternatively be expressed more explicitly using a two-stage formulation with a within-model and a between-model:

$$\begin{aligned} \mathbf{y}^*_{jk} &= \boldsymbol{\mu}_k + \boldsymbol{\Lambda}^{(2)}\boldsymbol{\eta}^{(2)}_{jk} + \boldsymbol{\epsilon}^{(2)}_{jk} \\ \boldsymbol{\mu}_k &= \boldsymbol{\mu} + \boldsymbol{\Lambda}^{(3)}\boldsymbol{\eta}^{(3)}_{k} + \boldsymbol{\epsilon}^{(3)}_{k}. \end{aligned} \quad (5)$$

The first equation for the latent responses $\mathbf{y}^*_{jk}$ represents a common factor model which includes random intercepts $\boldsymbol{\mu}_k$ that vary over clusters k. The second equation represents a common factor model for the random intercepts $\boldsymbol{\mu}_k$.

For the case of a single common factor at each level, a path diagram reflecting the above specification is given in Figure 2. Following the conventions used by Muthén and Muthén (2004), the models for the within and between covariance matrices are labeled 'within' and 'between'. The within-model shows the relationship between the observed responses and the common factor $\eta^{(2)}_1$ at the subject level. The solid circles attached to the responses indicate that the intercepts $\boldsymbol{\mu}_k$ of these responses vary randomly in the between-model. In the between-model, these random intercepts are shown as circles labeled with the names of the corresponding responses. These are modeled using a common and unique factors at the cluster level.

3.2.2 Reduced-form formulation

Substituting from the second line of (5) for $\boldsymbol{\mu}_k$ in the first line, we obtain the reduced form

$$\mathbf{y}^*_{jk} \;=\; \underbrace{\boldsymbol{\mu} + \boldsymbol{\Lambda}^{(3)}\boldsymbol{\eta}^{(3)}_k + \boldsymbol{\epsilon}^{(3)}_k}_{\boldsymbol{\mu}_k} + \boldsymbol{\Lambda}^{(2)}\boldsymbol{\eta}^{(2)}_{jk} + \boldsymbol{\epsilon}^{(2)}_{jk}.$$

A path diagram reflecting the reduced form is given in the left panel of Figure 3. Following the conventions in Rabe-Hesketh *et al.* (2004), nested frames represent the nested levels; variables located within the outer frame labeled 'cluster k' vary between clusters and have a k subscript and variables also inside the inner frame labeled 'unit j' vary between units within clusters and have both the j and k subscripts. Only common factors are enclosed in circles.

3.3 Variance components factor model

Instead of specifying separate factor models for the two levels, we could think of a single factor model defined for subjects in which the common factors have random intercepts varying between the clusters. In the unidimensional case, with a single common factor $\eta^{(2)}_{jk}$ at level 2, the measurement model for this factor is combined with a structural model of the form

$$\eta^{(2)}_{jk} \;=\; \eta^{(3)}_k + \zeta^{(2)}_{jk}. \tag{6}$$

Such a *variance components factor model* is analogous to a MIMIC ('Multiple-Indicator Multiple-Cause') model (e.g., Jöreskog & Goldberger, 1975) except that the common factor is not regressed on observed covariates but on a random intercept representing the effects of unobserved covariates at a higher level. An obvious application is in item response models if, for example, children's mean latent abilities vary randomly between schools (see e.g., Fox & Glas, 2001).

This model is a special case of the two-level factor model with the same number of common factors at both levels, no unique factors at level 3 and with factor loadings set equal across levels, $\boldsymbol{\Lambda}^{(2)} = \boldsymbol{\Lambda}^{(3)}$. Using the conventions of Muthén (e.g., Muthén and Muthén, 2004) the unidimensional variance components factor model would be depicted as in Figure 2 but without the short arrows in the 'between' model. Using our conventions, a natural representation is that given in Figure 3 (b).

The cluster-level unique factors in the two-level factor model can be thought of as representing differential item functioning between clusters. In Longford and Muthén's (1992) application to test scores in eight areas of mathematics for students nested in classes, the unique factors were interpretable as representing the variability in emphases between classrooms, partly due to tracking.

Note that if the factor loadings are set to 1 the model simply becomes a multilevel regression model. Such a model has been used by Raudenbush & Sampson (1999).

4 Multilevel Structural Equation Models

Just as for the single-level case, multilevel measurement models are sometimes of interest in their own right. However, it is often the nature of the relationships between latent variables at different levels that is the primary focus of the investigation.

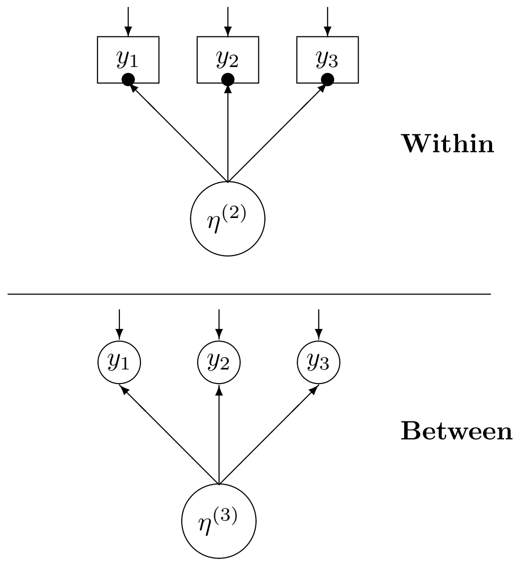

Figure 2: Path diagram of two-level factor model in within and between formulation

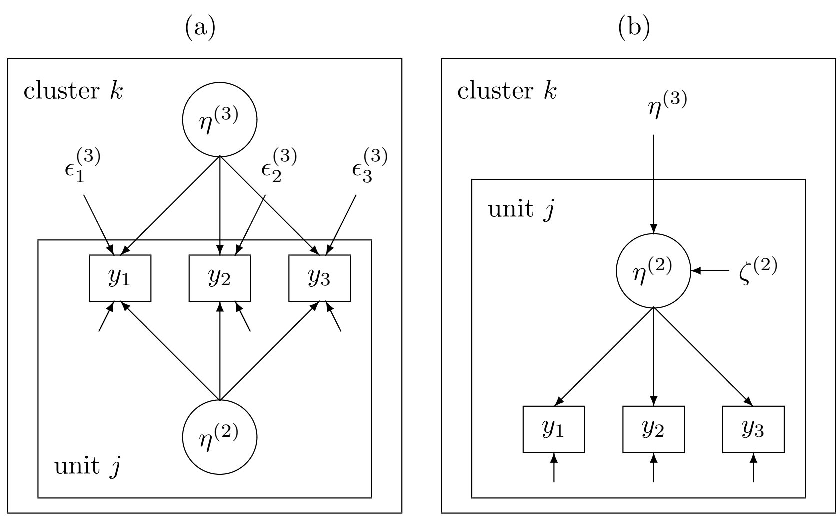

Figure 3: (a) general two-level factor model and (b) variance components factor model (Source: Skrondal and Rabe-Hesketh, 2004)

4.1 Single-level models

The M latent variables $\boldsymbol{\eta}_j$ are defined via a measurement model as described in Section 3. The structural model for the latent variables then allows these latent variables to be regressed on each other and on observed covariates. This model often has the form (e.g., Jöreskog, 1973; Muthén, 1984)

$$\boldsymbol{\eta}_j \;=\; \mathbf{B}\boldsymbol{\eta}_j + \boldsymbol{\Gamma}\mathbf{w}_j + \boldsymbol{\zeta}_j. \tag{7}$$

Here $\mathbf{B}$ is a regression parameter matrix for the relations among the latent variables $\boldsymbol{\eta}_j$, $\mathbf{w}_j$ is a vector of covariates, $\boldsymbol{\Gamma}$ is a parameter matrix for the regressions of the latent variables on the covariates, and $\boldsymbol{\zeta}_j$ is a vector of errors or disturbances. The relationships among the latent variables are recursive if the $\mathbf{B}$ matrix is strictly upper (or lower) diagonal.

4.2 Multilevel structural equation models

Multilevel structural equation models can be specified in a number of different ways. The most common approach is the traditional two-stage approach described for factor models in Section 3.2.1. In this case separate structural equation models are specified for the within and between covariance matrices (e.g., Muthén, 1994; Lee & Shi, 2001). A recent application of this approach in education is described by Everson and Millsap (2004). In contrast, the approach advocated here is based on including latent variables in random coefficient or generalized linear mixed models.

One possibility is to specify a conventional random coefficient model but let the *response variable* be a latent variable, for instance ability. The intercept and possibly effects of covariates are then specified as varying randomly between clusters (e.g., Fox and Glas, 2001). This is an extension of the unidimensional variance components factor model to include covariates and possibly random coefficients of covariates. The model includes direct paths from cluster-level latent variables to subject-level latent variables as shown for the variance components factor model in Figure 3b. While equivalent models can often be specified via separate models for the within and between covariance matrices, they require a large number of constraints, including nonlinear constraints (Rabe-Hesketh *et al.*, 2004). Furthermore, the simpler structure would not be apparent from separate diagrams for the within and between-models.

Remaining within the random coefficient framework, we can also let *covariates* be latent variables. If these covariates are cluster-specific, the model includes *responses varying at different levels*. This situation is accommodated within the framework suggested by Goldstein and McDonald (1988) and McDonald and Goldstein (1989) for continuous responses. Fox and Glas (2003) describe a model where both subject-level and cluster-level covariates are latent and where the measurement models are item response models. Unfortunately, the traditional two-stage formulations described in Section 3.2.1 cannot handle responses varying at different levels. This is a rather severe limitation for a multilevel structural equation model.

Rabe-Hesketh *et al.* (2004) develop the Generalized Linear Latent and Mixed Modeling (GLLAMM) framework consisting of a response model and a structural model. The response model has the form described in Section 3.1 but with

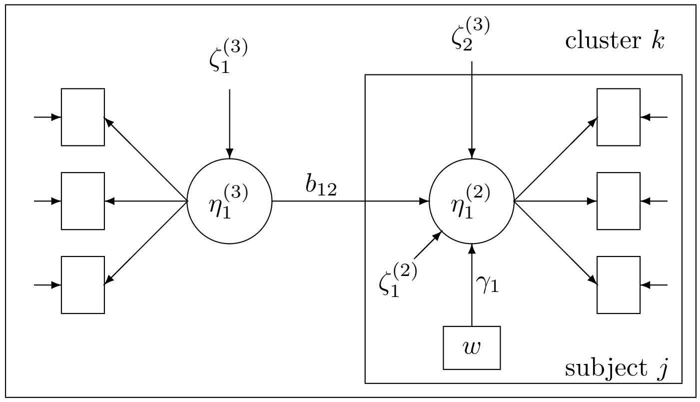

Figure 4: Multilevel structural equation model with latent dependent variable and latent covariate at level 2

L levels of nesting

$$\nu \;=\; \mathbf{X}\boldsymbol{\beta} + \sum_{l=2}^{L}\sum_{m=1}^{M_l} \boldsymbol{\eta}_m^{(l)}\mathbf{Z}_m^{(l)}\boldsymbol{\lambda}_m^{(l)} \tag{8}$$

where we have omitted the indices for units at different levels for notational simplicity. This model allows specification of random coefficient models, measurement models or both, as well as hybrid models. The structural model has the same form as (7) for single-level models but is specified for the vector $\boldsymbol{\eta}_j$ of all latent variables for subject j. This allows lower-level latent variables to be regressed on same or higher-level latent and observed variables. This framework permits specification of random coefficient models with latent responses or covariates at different levels. In addition, models in the two-stage formulation can also be specified. One limitation is that it is not possible to have a random coefficient of a latent covariate as this would correspond to a (cross-level) interaction among latent variables.

In Section 6 we will apply a model of the kind shown in Figure 4. A subject-level latent variable is regressed on a cluster-level latent variable and has a cluster-level random intercept. Moreover, the subject-level latent variable is regressed on covariates.

5 Estimation

5.1 Continuous responses

5.1.1 Maximum likelihood

For the continuous case, McDonald and Goldstein (1988, 1989) and Lee (1990) derived theory and succinct expressions for the likelihood, allowing two-level structural equation models to be estimated. For unbalanced multilevel designs with missing items, Longford and Muthén (1992) proposed a Fisher scoring algorithm whereas Raudenbush (1995) and Lee and Poon (1998) suggested EM algorithms.

5.1.2 Ad-hoc methods

Because these approaches require specialized software, several two-stage alternatives have been proposed. Muthén (1989) suggests an approach which corresponds to maximum likelihood for balanced data where all clusters have the same size n. In this case, the empirical covariance matrix $\mathbf{S}_W$ of the cluster-mean centered responses is a consistent and unbiased estimator for $\mathbf{\Sigma}_W$,

$$\mathrm{E}(\mathbf{S}_W) \;=\; \mathbf{\Sigma}_W.$$

In contrast, the expectation of the empirical covariance matrix $\mathbf{S}_B$ of the cluster means is

$$\mathrm{E}(\mathbf{S}_B) \;=\; \mathbf{\Sigma}_B + \frac{1}{n}\mathbf{\Sigma}_W.$$

Within and between structural equation models are specified for $\mathbf{\Sigma}_W$ and $\mathbf{\Sigma}_B$. Since $\mathbf{\Sigma}_W$ contributes to both $\mathrm{E}(\mathbf{S}_B)$ and $\mathrm{E}(\mathbf{S}_W)$, both models must be fitted jointly to the empirical covariance matrices $\mathbf{S}_B$ and $\mathbf{S}_W$. This can be accomplished by treating the two matrices as if they corresponded to different groups of subjects and performing two-group analysis with the required constraints. If there are only a relatively small number of different cluster sizes, a multiple group approach (with more than two groups) can be used to obtain maximum likelihood estimates. These approaches as well as an ad-hoc solution for the completely unbalanced case are described in detail in Muthén (1994) and Hox (2002).

Goldstein (1987, 2003) suggests using multivariate multilevel modeling to estimate $\mathbf{\Sigma}_W$ and $\mathbf{\Sigma}_B$ consistently by either maximum likelihood or restricted maximum likelihood. Structural equation models can then be fitted separately to each estimated matrix. Advantages of this approach are that unbalanced data and missing values are automatically accommodated, and that it is straightforward to extend to more hierarchical levels and to models where levels are crossed instead of nested.

An alternative ad-hoc approach, similar to the work by Korn and Whittemore (1979) was proposed by Chou *et al.* (2000). Here, a factor or structural equation model is estimated separately for each cluster. The estimates are subsequently treated as responses in a between-model, typically a regression model with between-cluster covariates and an unstructured multivariate residual covariance matrix. This approach allows, and indeed requires, all parameters to vary between clusters, including factor loadings.

A common feature of these two-stage procedures is that standard errors provided from the second stage are incorrect since they treat the output from the first stage as data or as empirical covariance matrices.

5.2 Non-continuous responses

For models with noncontinuous responses maximum likelihood estimation or Bayesian methods are typically used. Although computationally demanding, these methods automatically handle lack of balance and missing data and are straightforward to extend to include for instance mixed responses and nonlinear relations among latent variables. We note in passing that the ad-hoc approaches of Goldstein (1987, 2003) and Chou *et al.* (2000) discussed above can also be used for non-continuous responses.

5.2.1 Maximum likelihood estimation

The major challenge in maximum likelihood estimation of multilevel latent variable models for noncontinuous responses is to integrate out the latent variables since closed form results typically do not exist. Thus, integration usually proceeds by either by Monte Carlo simulation or using numerical methods.

Lee and Shi (2001) and Lee and Song (2004) use Monte Carlo EM (MCEM) algorithms, employing Gibbs sampling to evaluate the integrals in the E-step. Rabe-Hesketh *et al.* (2004) suggest using Newton-Raphson where the latent variables are integrated out using adaptive quadrature, see also Rabe-Hesketh *et al.* (2005).

5.2.2 Mean posterior estimation

As in other areas of statistics, Markov Chain Monte Carlo (MCMC) methods have recently attracted considerable interest in multilevel structural equation modeling. Interestingly, very diffuse priors are almost invariably specified in practice. The mean of the posterior distribution is in this case often quite close to the mode of the likelihood. MCMC can thus be viewed as a convenient and powerful way of implementing maximum likelihood estimation for complex models.

MCMC methods have been used by Ansari and Jedidi (2000), Fox and Glas (2001) and Goldstein and Browne (2005) for binary responses and by Song and Lee (2004) for continuous and ordinal responses.

6 Application: Student ability and teacher excellence

To investigate whether student ability measured at the student-level depends on teacher excellence measured at the school-level we analyze data from the Program for International Student Assessment (PISA) 2000 Assessment of Reading (OECD, 2001) using multilevel structural equation modeling.

6.1 Data description

The data consist of student responses to a reading test, a student background questionnaire and responses to a school questionnaire completed by principals.

At the student level, we focus on a unidimensional latent factor – the ability to interpret written information. We chose four items for this construct from the reading unit of the test. Three of the items have dichotomous responses and one item has ordinal responses. We included four observed covariates from the student questionnaire: Parents' education (one or both parents have higher education = 1, otherwise = 0), Male (male = 1, female = 0), Reading (some time spent on reading every day = 1, otherwise = 0), and English (English spoken at home = 1, otherwise = 0).

The school data include ordinal responses from principals to ten school survey questions measuring teacher excellence: teacher expectations, student-teacher relations, teacher turnover, teachers meeting individual students' needs, teacher absenteeism, teachers' strictness with students, teachers' morale, teachers' enthusiasm, teachers taking pride in the school, and teachers valuing academic achievement.

The sample comprises 2484 tenth grade students from 131 U.S. schools. School-level covariates were not included because this would have drastically reduced the number of schools due to missing data.

6.2 Model specification

In addition to developing measurement models for student interpretation ability and teacher excellence, we will estimate a structural equation model where student interpretation ability is regressed on student-level observed covariates and the school-level latent covariate teacher excellence. There is some doubt regarding the validity of the measurement of teacher excellence since this was based on a questionnaire completed by the principal. If teacher excellence is found to be predictive of student interpretation ability, this could be seen as supportive evidence for the validity of both instruments.

6.2.1 Student-level model

The single-level factor model discussed in Section 3.1 is estimated for the student data, where the common factor represents interpretation ability. The measurement model for interpretation ability $\eta^{(2)}_{1jk}$ can be written in terms of underlying continuous responses $\mathbf{y}^*_{jk}$. For item i for student j in school k we have

$$y^*_{ijk} \quad = \quad \beta_i + \lambda_i \eta^{(2)}_{1jk} + \epsilon_{ijk}, \qquad i=1,2,3,4. \tag{9}$$

Here β_i are item intercepts and λ_i factor loadings or discrimination parameters, and the ϵ_{ijk} have logistic distributions. Interpretation ability is measured by three dichotomous items and one ordinal item (item 2). For the dichotomous items ($i = 1$, 3 and 4),

$$y_{ijk} \;=\; \begin{cases} 1 & \text{if} \quad y^*_{ijk} > 0 \\ 0 & \text{otherwise} \end{cases},$$

and for the ordinal item ($i = 2$), the intercept β_2 is set to 0 and the threshold model is specified as

$$y_{2jk} \;=\; \begin{cases} 1 & \text{if} \quad -\infty \le y^*_{2jk} < \kappa_1 \\ 2 & \text{if} \quad \kappa_1 \le y^*_{2jk} < \kappa_2 \\ 3 & \text{if} \quad \kappa_2 \le y^*_{2jk} < \kappa_3 \\ 4 & \text{if} \quad \kappa_3 \le y^*_{2jk} < \infty. \end{cases}$$

This is a logistic graded response model (Samejima, 1969) where $-\beta_i$, $i=1,3,4$, can be interpreted as the thresholds for the dichotomous items.

We regress student interpretation ability on the four student background covariates (Parents' education, Male, Reading and English):

$$\eta^{(2)}_{1jk} \quad = \quad \boldsymbol{\gamma}' \mathbf{w}_{jk} + \zeta^{(2)}_{1jk}.$$

Here $\mathbf{w}_{jk} = [w_{1jk}, w_{2jk}, w_{3jk}, w_{4jk}]'$ is a vector of the four covariates, $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \gamma_3, \gamma_4]'$ the corresponding regression parameter vector and $\zeta^{(2)}_{1jk}$ a vector of student-level disturbances.

6.2.2 School-level model

Teacher excellence is measured by ten ordinal items, with response categories 'dissatisfied', 'somewhat satisfied' and 'satisfied'. The following model was used for the underlying continuous responses for items i and schools k:

$$y^*_{ik} \;=\; \eta^{(3)}_{1k} + \epsilon_{ik},$$

where $\eta^{(3)}_{1k}$ represents teacher excellence and ϵ_{ik} has a logistic distribution. The ordinal responses are generated from the threshold model

$$y_{ik} \;=\; \begin{cases} 1 & \text{if} \quad -\infty \leq y^*_{ik} < \alpha_1 + \tau_{i1} \\ 2 & \text{if} \quad \alpha_1 + \tau_{i1} \leq y^*_{ik} < \alpha_2 + \tau_{i2} \\ 3 & \text{if} \quad \alpha_2 + \tau_{i2} \leq y^*_{ik} < \infty, \end{cases}$$

where α_s $(s=1,2)$ is the sth threshold for item 1, whereas τ_{is} $(i=2,\ldots,10)$ is the difference in the sth threshold between item i and item 1. Thus, $\alpha_s+\tau_{is}$ corresponds to the threshold parameter κ_{is} for the ordinal responses as defined in Section 2.2.1. The model is a one-parameter version of the logistic graded response model which assumes that all items have the same discrimination.

The structural model is trivial if we do not wish to include school-level covariates:

$$\eta^{(3)}_{1k} \;=\; \zeta^{(3)}_{1k}.$$

6.2.3 Joint model

A joint model for the student data and school survey data combines the student-level and school-level models. In this example, students are the level-2 units and schools the level-3 units. Under the general response model in (8), the joint measurement model combines the item response model for the school survey data and a variance components factor model as discussed in Section 3.3 for the student data. A path diagram for this kind of model is shown in Figure 4.

For school k the response model is

$$\mathbf{y}^*_k \;=\; \mathbf{X}_k\boldsymbol{\beta} + \eta^{(2)}_{1jk}\mathbf{Z}^{(2)}_{1k}\boldsymbol{\lambda}^{(2)}_1 + \eta^{(3)}_{1k}\mathbf{Z}^{(3)}_{1k}\boldsymbol{\lambda}^{(3)}_1 + \eta^{(3)}_{2k}\mathbf{Z}^{(3)}_{2k}\boldsymbol{\lambda}^{(3)}_2.$$

Specifically, we can write the model for the responses of a student j from school k and a principal from school k as:

$$\begin{bmatrix} y^*_{1jk} \\ y^*_{2jk} \\ y^*_{3jk} \\ y^*_{4jk} \\ \hline y^*_{1k} \\ \vdots \\ y^*_{10,k} \end{bmatrix} \;=\; \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \hline 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_3 \\ \beta_4 \end{bmatrix} \;+\; \eta^{(2)}_{1jk} \begin{bmatrix} \mathbf{I}_{4\times 4} \\ \hline \mathbf{0}_{10\times 4} \end{bmatrix} \begin{bmatrix} 1 \\ \lambda^{(2)}_2 \\ \lambda^{(2)}_3 \\ \lambda^{(2)}_4 \end{bmatrix} \;+\; \eta^{(3)}_{1k} \begin{bmatrix} \mathbf{0}_{4\times 1} \\ \hline \mathbf{I}_{10\times 1} \end{bmatrix} 1$$

$$
+ \eta_{2k}^{(3)} \left[\mathbf{0}_{14\times 1} \right] 1 \; + \begin{bmatrix} \epsilon_{1jk} \\ \epsilon_{2jk} \\ \epsilon_{3jk} \\ \epsilon_{4jk} \\ \hline \epsilon_{1k} \\ \vdots \\ \epsilon_{10,k} \end{bmatrix}.
$$

In the vectors and matrices of the above model, student-level elements are placed above the horizontal lines and school-level elements below the horizontal lines. $\eta_{1jk}^{(2)}$ represents student interpretation ability, $\eta_{1k}^{(3)}$ teacher excellence, and $\eta_{2k}^{(3)}$ the school-level random intercept of interpretation ability. The latter is multiplied by zero for each item since the random intercept does not affect the items directly.

In the structural model, teacher excellence becomes an explanatory variable for interpretation ability. Moreover, interpretation ability is regressed on student-level covariates and the school-level random intercept $\eta_{2k}^{(3)}$ which allows students' mean ability to vary randomly between schools. The structural model can be written as

$$
\boldsymbol{\eta} \quad = \quad \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\mathbf{w} + \boldsymbol{\zeta}.
$$

Specifically,

$$
\begin{bmatrix} \eta_{1jk}^{(2)} \\ \eta_{1k}^{(3)} \\ \eta_{2k}^{(3)} \end{bmatrix} = \begin{bmatrix} 0 & b_{12} & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{1jk}^{(2)} \\ \eta_{1k}^{(3)} \\ \eta_{2k}^{(3)} \end{bmatrix} + \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_{1jk} \\ w_{2jk} \\ w_{3jk} \\ w_{4jk} \end{bmatrix} + \begin{bmatrix} \zeta_{1jk}^{(2)} \\ \zeta_{1k}^{(3)} \\ \zeta_{2k}^{(3)} \end{bmatrix}.
$$

In the structural model, the $\mathbf{B}$ matrix defines the relationship among the latent factors at different levels. In particular, the cross-level coefficient b_{12} represents the effect of school-level teacher excellence on student-level interpretation ability.

6.2.4 Results

Maximum likelihood estimates for the models considered above are given in Table 1. The estimates were obtained using *gllamm* (Rabe-Hesketh *et al.*, 2004) which uses adaptive quadrature (Rabe-Hesketh *et al.*, 2005) and runs in Stata (StataCorp, 2005).

In the student-level measurement model, the item difficulties for the dichotomous items are $-\beta_i/\lambda_i$. For the ordinal item, the κ_s represent the thresholds. Interpretation of these parameters is facilitated by inspecting the item characteristic curves shown in Figure 5.

Overall girls perform slightly better than boys as do students who read often or speak English at home. However, parents' education has a negligible estimated effect on student performance (not significant at the 5% level) which is somewhat surprising. This could be due to inaccurate reporting of students

Parameter		Student model		School model		Joint model	
		Est	(SE)	Est	(SE)	Est	(SE)
Student-level:							
β_1	[Item 1, intercept]	1.16	(0.16)			1.15	(0.16)
κ_1	[Item 2, threshold 1]	-0.22	(0.57)			-0.25	(0.58)
κ_2	[Item 2, threshold 2]	-0.92	(0.61)			0.91	(0.62)
κ_3	[Item 2, threshold 3]	2.15	(0.79)			2.16	(0.77)
β_3	[Item 3, intercept]	0.23	(0.25)			0.25	(0.23)
β_4	[Item 4, intercept]	-0.94	(0.23)			-0.94	(0.22)
λ_1	[Item 1, loading]	1				1	
λ_2	[Item 2, loading]	5.44	(2.88)			5.07	(2.38)
λ_3	[Item 3, loading]	2.05	(0.71)			1.80	(0.59)
λ_4	[Item 4, loading]	1.51	(0.54)			1.40	(0.46)
γ_1	[Parents' education]	-0.02	(0.05)			-0.02	(0.05)
γ_2	[Male]	-0.11	(0.06)			-0.12	(0.06)
γ_3	[Reading]	0.16	(0.08)			0.16	(0.08)
γ_4	[English]	0.27	(0.13)			0.30	(0.13)
$\text{var}(\zeta_1^{(2)})$	[Interpretation ability]	0.20	(0.12)			0.19	(0.10)
School-level:							
α_1	[Threshold parameters]			-1.86	(0.31)	-1.86	(0.31)
τ_{21}				-1.30	(0.47)	-1.30	(0.47)
τ_{31}				-0.84	(0.44)	-0.84	(0.43)
τ_{41}				0.39	(0.37)	0.39	(0.37)
τ_{51}				-0.41	(0.40)	-0.41	(0.40)
τ_{61}				-2.54	(0.67)	-2.54	(0.67)
τ_{71}				-0.33	(0.40)	-0.33	(0.40)
τ_{81}				-2.03	(0.56)	-2.03	(0.56)
τ_{91}				-2.29	(0.60)	-2.29	(0.60)
$\tau_{10,1}$				-2.30	(0.60)	-2.30	(0.60)
α_2				1.69	(0.31)	1.69	(0.31)
τ_{22}				0.06	(0.38)	0.06	(0.38)
τ_{32}				-0.56	(0.37)	-0.56	(0.37)
τ_{42}				0.86	(0.42)	0.86	(0.42)
τ_{52}				-0.36	(0.37)	-0.36	(0.37)
τ_{62}				-1.44	(0.36)	-1.43	(0.36)
τ_{72}				0.35	(0.40)	0.35	(0.40)
τ_{82}				0.29	(0.39)	0.29	(0.39)
τ_{92}				-0.67	(0.37)	-0.67	(0.37)
$\tau_{10,2}$				-1.79	(0.36)	-1.79	(0.36)
b_{12}	[Cross-level coefficient]					0.02	(0.03)
$\text{var}(\zeta_1^{(3)})$	[Teacher excellence]			2.19	(0.42)	2.19	(0.42)
$\text{var}(\zeta_2^{(3)})$	[Intercept]					0.05	(0.04)

Table 1: Maximum likelihood estimates for reading test data

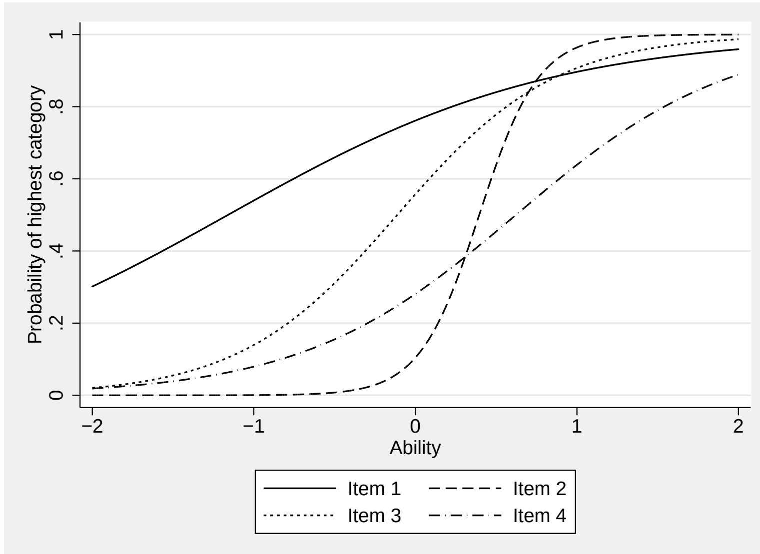

Figure 5: Item characteristic curves for the four interpretation items

on their parents' education or insufficient reliability for interpretation ability due to the small number of items.

The school-level model includes threshold parameters for the ten ordinal items. In category 2 (somewhat satisfied), the principals find "teachers' strictness with students" (item 6) the easiest to endorse and "teachers meeting individual students' needs" (item 4) the most difficult. In category 3 (satisfied), "teachers valuing academic achievement" (item 10) is the easiest and "teachers meeting individual students' needs" is once again the most difficult to endorse.

The school random intercept variance is estimated as .05 which is negligible for a logit model. The teacher excellence variance is estimated as 2.19. Somewhat surprisingly, the cross-level effect of teacher excellence on student interpretation ability appears to be negligible. One consequence of this is that the student-level and school-level parameters in the joint model do not differ much from those in the individual student and school models. The small estimated regression coefficient casts some doubt on the validity of the principal's assessment of teacher excellence based on the school questionnaire.

References

Ansari, A., & Jedidi, K. 2000. Bayesian factor analysis for multilevel binary observations. *Psychometrika*, **65**, 475–496.

Bartholomew, D. J. 1987. *Latent Variable Models and Factor Analysis*. Oxford: Oxford University Press.

Chou, C.-P., Bentler, P. M., & Pentz, M. A. 2000. Two-stage approach to multilevel structural equation models: Application to longitudinal data. *Pages 33–49 of:* Little, T. D., Schnabel, K. U., & Baumert, J. (eds), *Modeling*

Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Specific Examples. Mahwah, NJ: Erlbaum.

De Boeck, P., & Wilson, M. 2004. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York: Springer.

Everson, H. T., & Millsap, R. E. 2004. Beyond individual differences: Exploring school effets on SAT scores. *Educational Psychologist*, **39**, 157–172.

Fox, J. P., & Glas, C. A. W. 2001. Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, **66**, 271–288.

Fox, J. P., & Glas, C. A. W. 2003. Bayesian modeling of measurement error in predictor variables using item response theory. *Psychometrika*, **68**, 169–191.

Goldstein, H. 1987. Multilevel covariance component models. *Biometrika*, **74**, 430–431.

Goldstein, H. 2003. *Multilevel Statistical Models (Third Edition)*. London: Arnold.

Goldstein, H., & Browne, W. J. 2005. Multilevel factor analysis models for continuous and discrete data. *Pages 453–475 of:* Maydeu-Olivares, A., & McArdle, J.J. (eds), *Contemporary Psychometrics: A Festschrift for Roderick P. McDonald*. Mahwah, NJ: Erlbaum.

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

Hox, J. 2002. *Multilevel Analysis: Techniques and Applications*. Mahwah, NJ: Erlbaum.

Jöreskog, K. G. 1973. A general method for estimating a linear structural equation system. *Pages 85–112 of:* Goldberger, A. S., & Duncan, O. D. (eds), *Structural Equation Models in the Social Sciences*. New York: Seminar.

Jöreskog, K. G., & Goldberger, A. S. 1975. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, **70**, 631–639.

Korn, E. L., & Whittmore, A. S. 1979. Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics*, **35**, 795–804.

Lee, S.-Y. 1990. Multilevel analysis of structural equation models. *Biometrika*, **77**, 763–772.

Lee, S.-Y., & Shi, J.-Q. 2001. Maximum likelihood estimation of two-level latent variable models with mixed continuous and polytomous data. *Biometrics*, **57**, 787–794.

Lee, S.-Y., & Song, X.-Y. 2004. Maximum likelihood analysis of a general latent variable model with hierarchically mixed data. *Biometrics*, **60**, 624–636.

Linda, N. Y., Lee, S.-Y., & Poon, W.-Y. 1993. Covariance structure analysis with three level data. *Computational Statistics & Data Analysis*, **15**, 159–178.

Longford, N. T. 1993. *Random Coefficient Models*. Oxford: Oxford University Press.

Longford, N. T., & Muthén, B. O. 1992. Factor analysis for clustered observations. *Psychometrika*, **57**, 581–597.

McDonald, R. P., & Goldstein, H. 1989. Balanced and unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology*, **42**, 215–232.

Mellenbergh, G. J. 1994. Generalized linear item response theory. *Psychological Bulletin*, **115**, 300–307.

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

Muthén, B. O. 1989. Latent variable modeling in heterogeneous populations. *Psychometrika*, **54**, 557–585.

Muthén, B. O. 1994. Multilevel covariance structure analysis. *Sociological Methods & Research*, **22**, 376–398.

Muthén, L. K., & Muthén, B. O. 2004. *Mplus User's Guide (Third Edition)*. Los Angeles, CA: Muthén & Muthén.

OECD. 2001. *Knowledge and Skills for Life: First Results from PISA 2000*. Paris: OECD.

Poon, W. Y. Lee S. Y. 1998. Analysis of two-level structural equation models via EM-type algorithm. *Statistica Sinica*, **8**, 749–766.

Poon, W. Y., & Lee, S. Y. 1992. Maximum likelihood and generalized least squares analyses of two-level structural equation models. *Statistics and Probability Letters*, **14**, 25–30.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. 2004a. Generalized multilevel structural equation modeling. *Psychometrika*, **69**, 167–190.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. 2004b. *GLLAMM Manual*. Tech. rept. 160. U.C. Berkeley Division of Biostatistics Working Paper Series. Downloadable from http://www.bepress.com/ucbbiostat/paper160/.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. 2005. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, **128**, 301–323.

Raudenbush, S. W. 1995. Maximum likelihood estimation for unbalanced multilevel covariance structure models via the EM algorithm. *British Journal of Mathematical and Statistical Psychology*, **48**, 359–370.

Raudenbush, S. W., & Sampson, R. 1999. Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Pages 1–41 of:* Marsden, P. V. (ed), *Sociological Methodology 1999*. Oxford: Blackwell.

Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. 2003. A nonlinear mixed model framework for item response theory. *Psychological Methods*, **8**, 185–205.

Samejima, F. 1969. *Estimation of Latent Trait Ability Using A Response Pattern of Graded Scores.* Bowling Green, OH: Psychometric Monograph 17, Psychometric Society.

Skrondal, A. 1996. *Latent Trait, Multilevel and Repeated Measurement Modelling with Incomplete Data of Mixed Measurement Levels.* Oslo: Section of Medical Statistics, University of Oslo.

Skrondal, A., & Rabe-Hesketh, S. 2003. Multilevel logistic regression for polytomous data and rankings. *Psychometrika*, **68**, 267–287.

Skrondal, A., & Rabe-Hesketh, S. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models.* Boca Raton, FL: Chapman & Hall/CRC.

Song, X.Y. & Lee, S.Y. 2004. Bayesian analysis of two-level structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, **57**, 29-52.

StataCorp. 2005. *Stata Statistical Software: Release 9.0.* College Station, TX.

Takane, Y., & de Leeuw, J. 1987. On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, **52**, 393–408.