Multilevel Mixture Factor Models

Roberta Varriale and Jeroen K. Vermunt
Department of Methodology and Statistics, Tilburg University

Factor analysis is a statistical method for describing the associations among sets of observed variables in terms of a small number of underlying continuous latent variables. Various authors have proposed multilevel extensions of the factor model for the analysis of data sets with a hierarchical structure. These Multilevel Factor Models (MFMs) have in common that – as in multilevel regression analysis – variation at the higher level is modeled using continuous random effects (or continuous latent variables). In this paper, we present an alternative multilevel extension of factor analysis which we call the Multilevel Mixture Factor Model (MFMM). It is based on the assumption that higher-level units belong to latent classes that differ in terms of the parameters of the factor model specified for the lower-level units. We demonstrate the added value of MMFM compared to MFM, both from a theoretical and applied perspective, and we illustrate the complementarity of the two approaches with an empirical application.

1 Introduction

Factor analysis (FA) is a commonly used statistical method for investigating the dimensionality of sets of observed responses, or, more technically, for describing the associations among sets of manifest variables, referred to as indicators or items, in terms of a smaller number of underlying continuous latent variables, referred to as factors (Bartholomew

& Knott, 1999). Whereas the term "factor model" has been traditionally reserved for models in which the indicators are continuous variables, in the more recent FA literature, it is also used to refer to models for dichotomous or ordinal indicators. Here, we use this broader definition, which means that we not only deal with classical FA but also with models which are known as (multidimensional) item response theory (IRT) models.

As most statistical methods, standard FA is based on the assumption that the available data set is a sample consisting of independent observations. However, this assumption is inadequate with hierarchical or multilevel data; that is, when the indicators of the factor model are measured on individuals (lower-level or micro-level units) which are nested within groups (higher-level or macro-level units) sharing common environments, experiences, and interactions. Examples are data on students nested within schools, patients nested within hospitals, employees nested within organizations, citizens nested within regions, etc. In such situations, it is more appropriate to use multilevel techniques which not only account for the dependencies between observations due to the hierarchical data structure, but also make it possible to determine to what extent the phenomenon under study can be explained by group- or macro-level factors (Goldstein, 2003; Hox, 2002; Snijders & Bosker, 1999). Various authors proposed multilevel extensions of standard FA and IRT (see, e.g., Fox & Glas, 2001; Goldstein & McDonald, 1988; Grilli & Rampichini, 2007; Longford & Muthén, 1992; Muthén, 1991; Muthén & Satorra, 1989). As in multilevel regression models, the basic idea of these Multilevel Factor Models (MFMs) is that some of the model parameters are allowed to randomly vary across groups. These random effects are, in fact, continuous latent variables (or factors) at the group level. This means that MFMs assume the presence of latent factors at both the lower and the higher level of the analysis. As standard factor models, these models can be used in either an exploratory or confirmatory manner.

In this paper, we present an alternative and complementary approach for FA with multilevel data. More specifically, we propose modeling between group heterogeneity using a mixture model which assumes that the higher-level units belong to latent classes which differ in the parameter values of the factor model defined for the lower-level units. This Multilevel Mixture Factor Model (MMFM) yields a clustering of groups based on the factor model parameters, which in many applications is a more natural way to describe group differences than the typical multilevel variance decomposition. Examples of applications include clustering of schools based on the performance of students, clustering of organizations based on the satisfaction of their clients, and clustering of regions based on the political opinions of their citizens. Another advantage compared to the MFM is that the MMFM is more robust in the sense that it does not introduce possible inappropriate and unverifiable assumptions about the distribution of the higher-level variation (Aitkin, 1999). A third advantage is that it is computational less demanding than the MFM, especially when applied with discrete items.

The proposed MMFM is related to two other extensions of standard FA and IRT; that is, to multi-group FA/IRT and mixture FA/IRT. Also in a multiple-group FA model parameters are allowed to differ across groups (Bollen, 1989; Meredith, 1993; Muthén, 1989). More specifically, for parameters which vary across groups one obtains a separate estimate for each group. In fact, such a model can be seen as a MMFM in which there are as many latent classes as groups and where each group belongs to a different class. It will be clear that such an approach becomes problematic when the number of groups is large, in which case it is attractive to cluster the groups into a relatively small number of latent classes. The MMFM can thus be seen as a tool for multiple-group FA that can also be used when the number of groups is large.

In mixture FA and IRT (Maij-de Meij, Kelderman & van der Flier, 2008; McLachlan &

Peel, 2000; Lubke & Muthén, 2005; Rost, 1990; Yung, 1997), individuals are assumed to belong to latent classes which differ with respect to the parameters of the specified measurement model. In fact, these are multiple-group FA/IRT models in which the grouping is unobserved. The difference between mixture FA/IRT and the proposed MMFM is that in the former the mixture is at the individual level whereas in the latter it is at the higher-level of the analysis. Despite this important difference, there is an interesting connection between the two: Stating that higher-level units belong to latent classes is equivalent to stating that individuals belonging to the same higher-level unit belong to the same latent class (Asparouhov & Muthén, 2008). In other words, the MMFM can be seen as a mixture factor model with a specific structure for the individual latent class memberships, a structure derived from the hierarchical data structure.

[INSERT TABLE 1 ABOUT HERE]

The idea of using a discrete mixture distribution at the higher level is similar to what Vermunt (2003, 2008a) proposed in the context of latent class analysis. More specifically, he presented two types of multilevel latent class models: one in which group-level heterogeneity is modeled using continuous random effects and one in which it is modeled by assuming that not only individuals but also groups belong to latent classes. Table 1 provides Vermunt's (2008a) four-fold classification of two-level latent variables models based on the specification for the lower- and higher-level latent variables. As can be seen, the MFM corresponds to the situation with continuous latent variables at both levels (IV), and the MMFM proposed in this paper to the situation with continuous latent variables at the lower level and discrete latent variables at the higher level (III). The other two types with discrete latent variables at the lower level (I and II) are the two types of multilevel latent class models introduced by Vermunt (2003). Another example of a type II model is the two-level mixture regression model proposed by Muthén and Asparouhov (2009).

The MMFM fits naturally into the generalized latent variable modeling framework described among others by Skrondal and Rabe-Hesketh (2004), Muthén and Asparouhov (2008), and Vermunt (2008b), and implemented in the GLLAMM (Rabe-Hesketh, Skrondal & Pickles, 2004), Mplus (Muthén & Muthén, 1998-2007), and Latent GOLD (Vermunt & Magidson, 2005, 2008) software packages. Using this framework, related multilevel latent variable models with a mixture at the higher level and one or more continuous latent variables at the lower level have been proposed by among others Muthén & Asparouhov (2008), Palardy & Vermunt (in press), and Vermunt (2008b).

The remainder of this paper is organized as follows. The next section describes the standard factor model, as well as the MFM and the MMFM. Subsequently, we discuss estimation using maximum likelihood and model selection issues. Then, we illustrate the use of the MFM and MMFM by means of an empirical example on students' satisfaction with various aspects of the university. The last section presents the main conclusions.

2 Model specification

2.1 The factor model

Let y_{hi} denote the observed response of individual i $(i = 1, \ldots, N)$ on indicator h $(h = 1, \ldots, H)$, and η_{mi} the unobserved score of individual i on common factor m $(m = 1, \ldots, M)$, where N, H, and M are the total number of individuals, items, and factors. The vectors of responses and factor scores of individual i are denoted by $\mathbf{y}_i = (y_{1i}..., y_{Hi})'$ and $\boldsymbol{\eta}_i = (\eta_{1i}, \ldots, \eta_{Mi})'$, respectively. In FA, a series of H regression models are used to define the relationships between the latent variables $\boldsymbol{\eta}_i$ and the indicators y_{hi}. To accommodate for the various possible scale types of the indicators, we use response models from the generalized linear modeling family, which are specified via a linear predictor, a link function, and an error distribution from the exponential family (Skrondal & Rabe-Hesketh, 2004).

We refer to the linear predictor of the response model for indicator h by υ_{hi}. In a factor analytic model the linear predictor has the following form:

$$\upsilon_{hi} = \mu_h + \sum_{m=1}^{M} \lambda_{mh}\eta_{mi}, \tag{1}$$

where μ_h is an item intercept and λ_{mh} a factor loading. The linear predictor is connected to y_{hi} as follows:

$$g(E(y_{hi}|\boldsymbol{\eta}_i)) = \upsilon_{hi}, \tag{2}$$

This equation shows that after applying an appropriate transformation $g(\cdot)$, which is referred to as the link function, the expected value of y_{hi} conditional on the latent factors equals the linear predictor. The choice of the link function depends on the scale type of the indicators. For continuous responses, one usually uses an identity link – $E(y_{hi}|\boldsymbol{\eta}_i) = \upsilon_{hi}$ – whereas for binary responses one may use among others a logit link – $\log E(y_{hi}|\boldsymbol{\eta}_i)/[1 - E(y_{hi}|\boldsymbol{\eta}_i)] = \upsilon_{hi}$, implying that $E(y_{hi}|\boldsymbol{\eta}_i) = \exp(\upsilon_{hi})/[1 + \exp(\upsilon_{hi})]$.

The definition of the H response models is completed by the specification of the distribution of the indicators' residuals $e_{hi} = y_{hi} - E(y_{hi}|\boldsymbol{\eta}_i)$ or, equivalently, of the conditional density of y_{hi} given the latent variables $f(y_{hi}|\boldsymbol{\eta}_i)$. The typical choice in standard FA with continuous responses is the normal distribution; that is, $f(e_{hi}) = f(y_{hi}|\boldsymbol{\eta}_i) \sim N(0, \sigma_h^2)$. With dichotomous responses, $f(y_{hi}|\boldsymbol{\eta}_i)$ is usually assumed be a Bernoulli distribution.

It should be noted that it is assumed that conditionally on the latent variables $\boldsymbol{\eta}_i$, the H observed responses are independent of each other, an assumption that is referred to as the local independence assumption (Bartholomew & Knott, 1999). It implies that the joint distribution of $f(\mathbf{y}_i|\boldsymbol{\eta}_i)$ can be written as a product of the marginal distributions $f(y_{hi}|\boldsymbol{\eta}_i)$; that is, $f(\mathbf{y}_i|\boldsymbol{\eta}_i) = \prod_{h=1}^{H} f(y_{hi}|\boldsymbol{\eta}_i)$.

For FA with ordinal indicators several types of specifications of the response model are available. One of the most used is the proportional odds model that uses a cumulative logit link function. Let s be a particular category and S the total number categories of the ordinal indicator y_{hi} ($s = 1, \ldots, S$), and let $P(y_{hi} \leq s|\boldsymbol{\eta}_i)$ be the probability of responding in category s or lower conditional on the common factors. The proportional odds model can now be expressed as follows:

$$\log \frac{P(y_{hi} \leq s|\boldsymbol{\eta}_i)}{1 - P(y_{hi} \leq s|\boldsymbol{\eta}_i)} = \alpha_{hs} - v_{hi} \quad s = 1, \ldots, S-1, \tag{3}$$

where the α_{hs} – with $\alpha_{h1} < \ldots < \alpha_{hS-1}$ – are threshold parameters to be estimated. An important feature of this model is that the parameters involved in v_{hi} are invariant for how the ordinal categories of y_{hi} are constructed. Other possible choices for the link function include the probit and complementary log-log links.

The specification of the factor model is completed with the specification of the distribution of the common factors. In FA, it is usually assumed that the common factors come from a multivariate normal distribution; that is,

$$\boldsymbol{\eta}_i \sim MN(\mathbf{0}, \boldsymbol{\Psi}), \tag{4}$$

where $\boldsymbol{\Psi}$ is an $M \times M$ covariance matrix with elements $\psi_{mm'}$.

2.2 The multilevel factor model

The standard factor model assumes that the indicators are measured on a set of independently sampled units, an assumption that is violated when individuals are nested within groups sharing common environments, experiences and interactions. But fortunately it is possible to define factor models which can deal with hierachical data structures. As in other types of models for multilevel data, a multilevel factor model (MFM) is obtained by allowing that higher-level units differ randomly with respect to particular model parameters. In the factor model these are intercepts, thresholds, residual variances, factor loadings, factor means, and factor (co)variances. Whereas we limit ourselves to the situation that there are two hierarchical levels, the extension to three and more levels is conceptually straightforward. Let y_{hij} denote the observed response on indicator h ($h = 1, \ldots, H$) of individual i ($i = 1, \ldots, n_j$) within group j ($j = 1, \ldots, J$). Note that J denotes the number of groups and n_j the number of individuals within group j, a number that may vary across groups. The total number of individuals is $N = \sum_{j=1}^{J} n_j$. A MFM can either be seen as a two-level model for multivariate responses, with individuals i being the level-1 units and groups j the level-2 units, or as a three-level model for univariate responses, with indicators h being the level-1 units, individuals i being the level-2 units, and groups j the level-3 units. As Skrondal and Rabe-Hesketh (2004), we will formulate the MFM as a three-level model where subscripts and superscripts 1, 2, and 3 are used to refer to quantities relating to indicators, individuals, and groups, respectively.

The linear predictor in the response model for y_{hij} is denoted by v_{hij}; $\boldsymbol{\eta}_{ij}^{(2)} = (\eta_{1ij}^{(2)}, \ldots, \eta_{M_2ij}^{(2)})'$ contains the M_2 common factors varying at the individual level and $\boldsymbol{\eta}_{j}^{(3)} = (\eta_{1j}^{(3)}, \ldots, \eta_{M_3j}^{(3)})'$ the M_3 common factors varying at the group level. The MFM is expressed by following three sets of regression equations:

$$v_{hij} = \mu_{hj} + \sum_{m_2=1}^{M_2} \lambda_{m_2h}^{(2)} \eta_{m_2ij}^{(2)} \tag{5}$$

$$\mu_{hj} = \mu_h + \sum_{m_3=1}^{M_3} \lambda_{m_3h}^{(3)} \eta_{m_3j}^{(3)} + e_{hj}^{(3)} \tag{6}$$

$$\eta_{m_2ij}^{(2)} = \sum_{m_3=1}^{M_3} \beta_{m_3m_2}^{(3)} \eta_{m_3j}^{(3)} + e_{m_2ij}^{(2)}. \tag{7}$$

The first of these equations is very similar to the standard factor model presented in equation (1). The main modification compared to the standard FA equation is that the linear predictor, the item intercept, and the factors have an additional subscript j

indicating that their values depend on the group. The other changes are just notational to be able to distinguish level-2 quantities from level-3 quantities: M and m have a subscript 2, and λ and η have a superscript (2).

Equations (6) and (7) show the two roles that the group-level factor $\eta^{(3)}_{m_3j}$ may play in a MFM. Equation (6) shows that these may directly affect the indicators. In this equation, μ_h is the mean intercept of item h, and $\lambda^{(3)}_{m_3h}$ and $e^{(3)}_{hj}$ are, respectively, the factor loadings and the item-specific errors (or unique factors) at the highest level of the analysis. Equation (7) shows that the higher-level factors may also affect the factor at the lower level. In this equation, $\beta^{(3)}_{m_3m_2}$ is the coefficient (structural parameter) "linking" the latent variable $\eta^{(3)}_{m_3j}$ to the latent variable $\eta^{(2)}_{m_2ij}$ and $e^{(2)}_{m_2ij}$ is a residual term.

As in a one-level factor model, the conditional expectation of the response y_{hij} given the latent variables at different levels is "linked" to the linear predictor v_{hij} via a link function:

$$g(E(y_{hij}|\boldsymbol{\eta}^{(2)}_{ij}, \boldsymbol{\eta}^{(3)}_{j})) = v_{hij}, \tag{8}$$

where the full model for v_{hij} is obtained by substituting equations (6) and (7) into equation (8). Again, different types of distributional forms can be used for the residuals of the indicators (within higher-level units).

Similarly to standard factor models, it is assumed that

$$\mathbf{e}^{(2)}_{ij} = \boldsymbol{\eta}^{(2)}_{ij}|\boldsymbol{\eta}^{(3)}_{j} \sim MN(\mathbf{0}, \boldsymbol{\Psi}^{(2)}) \tag{9}$$

$$\boldsymbol{\eta}^{(3)}_{j} \sim MN(\mathbf{0}, \boldsymbol{\Psi}^{(3)}) \tag{10}$$

$$\mathbf{e}^{(3)}_{j} \sim MN(\mathbf{0}, \boldsymbol{\Omega}^{(3)}), \tag{11}$$

where $\boldsymbol{\Psi}^{(2)}$ and $\boldsymbol{\Psi}^{(3)}$ represent $M_2 \times M_2$ and $M_3 \times M_3$ covariance matrices with elements $\psi^{(2)}_{mm'}$ and $\psi^{(3)}_{mm'}$, respectively, and $\boldsymbol{\Omega}^{(3)}$ an $H \times H$ covariance matrix with elements $\omega^{(3)}_{hh'}$. Typically, because of the local independence assumption, $\boldsymbol{\Omega}^{(3)}$ is assumed to be diagonal.

[INSERT FIGURE 1 ABOUT HERE]

Figure 1(a) depicts the MFM. Following the conventions, circles represent latent variables and rectangles observed variables, arrows connecting latent and/or observed variables represent direct effects, which do not need to be linear, and dotted lines represent correlations among latent variables or among items. A nesting of frames is used to express the hierarchical levels; that is, the outer frame contains variables varying between groups and the inner frame variables varying between individuals within groups (Skrondal & Rabe-Hesketh, 2004).

Whereas above we defined the MFM in its most general form, in most applications one will use specific restricted versions of this model. One interesting special case is obtained with $M_3 = M_2 = M$, $\lambda^{(3)}_{m_3h} = 0$ for all m_3 and h, $\beta^{(3)}_{m_3m_2} = 0$ for $m_3 \neq m_2$, and for identification purposes $\beta^{(3)}_{m_3m_2} = 1$ for $m_3 = m_2$; that is, by assuming the group-level factor $\eta^{(3)}_{mj}$ affects the response only indirectly via $\eta^{(2)}_{mij}$. Such a specification is most useful if one wishes to model the latent structure at the individual level, while accounting for the multilevel data structure by allowing for between-group variation in the means of the individual-level latent factors. With the additional restriction $\omega^{(3)}_{hh'} = 0$ for $h = 1, \ldots, H$, indicating that there are no unique factors $e^{(3)}_{hj}$ at the group level, one obtains the so-called variance component factor model (Skrondal & Rabe-Hesketh, 2004). Also the multilevel IRT model proposed by Fox and Glas (2001) is of this form, but expanded with covariates in the model for $\eta^{(2)}_{mij}$. The variance component factor model, which is represented in Figure 1(b) allows splitting the variance of each factor into level-specific components. Note that this model assumes that the factor structure is the same at both levels, and it can therefore also be obtained by setting $\lambda^{(3)}_{m_3h} = \lambda^{(2)}_{m_2h}$ for $m_2 = m_3$ and $\beta^{(3)}_{m_3m_2} = 0$ for all m_3 and m_2.

An example of an application in which the variance component factor model may be of interest is an educational study focusing on how pupils abilities vary across schools. In such an application, the factor model is used to measure the students abilities affecting the performance on a set of test items. Since students attending the same school are taught by the same teachers, come from the same neighborhood, and so on, their abilities may be more similar than those of students attending different schools. The variance component factor model allows modeling these within-school dependencies by letting the averages of the students' abilities vary across schools. This yields an estimate of the relative importance of school effects on the students' abilities.

Another interesting special case of the MFM described above is obtained with the restriction $\beta^{(3)}_{m_3 m_2} = 0$, which means that the latent variables at the higher level $\eta^{(3)}_{m_3 j}$ do not affect the latent variables at the lower level $\eta^{(2)}_{m_2 ij}$. Instead, the indicators are directly affected by both $\eta^{(2)}_{m_2 ij}$ and $\eta^{(3)}_{m_3 j}$, where it is important to note that the number of factors as well as the factor loadings may differ across levels. The latter implies that the model allows for completely different factor structures at the lower and higher level of analysis. This model is typically used in more exploratory manner, as proposed among others by Goldstein and McDonald (1988), Grilli and Rampichini (2007), and Longford and Muthén (1992).

Following the educational example, it may be that the way the responses on the test items vary between schools is different from the way these vary within schools. More specifically, it could be that at the student level the responses are affected by two factors - say mathematical and reading ability – whereas at the school level they are affected a single general ability factor. In such a case, the second special case of the MFM should be used instead of the variance components model.

2.3 The multilevel mixture factor model

We will now describe an alternative approach to FA with multilevel data sets. Rather than modeling the between group heterogeneity using group-level continuous factors or random effects, we propose modeling group differences by assuming that these belong to one of K latent classes or mixture components. We call the resulting model a Multilevel Mixture Factor Model (MMFM). Similar mixture model extensions have been proposed for FA and IRT models for non-hierachical data (see, e.g. Maij-de Meij et al., 2008; McLachlan & Peel, 2000; Lubke & Muthén, 2005; Rost, 1990; Yung, 1997). The motivation was always that one wished to relax the assumption of the standard factor model that the sample originated from a single homogeneous population. Finite mixture and latent class models (Goodman, 1974; Lazarsfeld & Henry, 1968; McLachlan & Peel, 2000) are designed to check this assumption and to examine whether there is evidence for the existence of unobserved subpopulations (latent classes).

The main difference between our MMFM and existing mixture FA and IRT models is that we postulate a discrete mixture distribution at the group instead of the individual level; that is, we use a discrete latent variable to deal with the fact that groups do not originate from a single homogeneous population of groups as far as the FA parameters is concerned. The key difference with the MFMs discussed above is that these assume that all groups are different – each group has its own factor scores – whereas the MMFM assumes that groups can be classified into homogeneous classes. The latter is especially useful if the final aim of the analysis is to classify groups. However, the discrete specification can also be used as a way to approximate continuous higher-level variation without making strong assumptions about the distribution of the group-level factors. Such a semi- or non-parametric specification of the random effects is similar to what Aitkin (1999) and Vermunt and Van Dijk (2001) described for multilevel regression analysis (see also Skrondal & Rabe-Hesket, 2004).

Let K be the number of latent classes and k a particular latent class ($k = 1, ..., K$). The class membership of a group is represented using K indicator variables $\eta^{(3)}_{kj}$ taking on value 1 if group j belongs to latent class k and 0 otherwise. The vector of indicator variables is $\boldsymbol{\eta}^{(3)}_j = (\eta^{(3)}_{1j}, \ldots, \eta^{(3)}_{Kj})$. The MMFM can be expressed by using the following three regression equations:

$$\upsilon_{hij} = \mu_{hj} + \sum_{m_2=1}^{M_2} \lambda^{(2)}_{m_2h} \eta^{(2)}_{m_2ij} \quad (12)$$

$$\mu_{hj} = \mu_h + \sum_{k=1}^{K} \lambda^{(3)}_{kh} \eta^{(3)}_{kj} + e^{(3)}_{hj} \quad (13)$$

$$\eta^{(2)}_{m_2ij} = \sum_{k=1}^{K} \beta^{(3)}_{km_2} \eta^{(3)}_{kj} + e^{(2)}_{m_2ij}. \quad (14)$$

Note that these equations are very much the same as equations (5)-(7) defining the MFM. As before, the first equation is similar to the equation of a standard one-level factor model and the other two equations show how one obtains the relevant multilevel extension. The difference with the MFM is that the sums in the second and third equation are now over K latent indicators instead of M_3 higher-level continuous factors. As always with categorical variables, for identification purposes, one constraint has to be imposed per parameter set; that is, on $\lambda^{(3)}_{kh}$ for each h and on $\beta^{(3)}_{km_2}$ for each m_2; for example, $\sum_{k=1}^{K} \lambda^{(3)}_{kh} = 0$ or $\lambda^{(3)}_{1h} = 0$ and $\sum_{k=1}^{K} \beta^{(3)}_{km_2} = 0$ or $\beta^{(3)}_{1m_2} = 0$. Figure 2 depicts the MMFM, where the categorical latent variable is represented by a filled circle.

[INSERT FIGURE 2 ABOUT HERE]

The discrete latent variable $\boldsymbol{\eta}^{(3)}_j$ has a multinomial distribution, with:

$$\pi_k = P(\eta^{(3)}_{kj} = 1) = \frac{\exp(\gamma_k)}{\sum_{t=1}^{K} \exp(\gamma_t)}, \quad (15)$$

where $\sum_{k=1}^{K} \pi_k = 1$. The term γ_k in equation (15) represents the intercept term in the linear predictor of the logit model for the latent class probabilities (π_k). Models with covariates affecting the class membership probability are obtained by adding the relevant covariate effects to this linear term.

As for the MFM, two special cases of the MMFM are obtained by constraining either the $\lambda_{kh}^{(3)}$ in equation (13) or the $\beta_{km_2}^{(3)}$ in equation (14) to 0. The model with $\lambda_{kh}^{(3)} = 0$ and $\beta_{km_2}^{(3)}$ estimated freely assumes that group-level latent classes do not differ in their responses given possible differences in mean abilities. Following the educational example described before, such a specification is most useful if one wishes to classify schools based on their students' abilities. The specification with $\beta_{km_2}^{(3)} = 0$ and $\lambda_{kh}^{(3)}$ estimated freely implies that latent classes differ with respect to the items' intercepts μ_{hj} but not with respect to $\eta_{m_2ij}^{(2)}$. In the educational example, such model would be useful if the researcher wishes to classify schools based on the means of the separate indicators. The factor model at the lower level would then be of less interest, and just a way to model dependencies between responses within subjects.

3 Estimation and model selection issues

3.1 Likelihood, parameter estimation, and posterior analysis

The unknown parameters of the MFM and MMFM described in this paper can be estimated by means of maximum likelihood. This involves maximizing the following marginal likelihood function:

$$L = \prod_{j=1}^{J} f(\mathbf{y}_j), \tag{16}$$

where $\mathbf{y}_j = (\mathbf{y}_{1j}, ...\mathbf{y}_{n_jj})$ is the vector containing all observed responses of group j, and $f(\mathbf{y}_j)$ is the probability density of these observations. As can be seen the marginal likelihood is obtained as a product of the likelihood contributions of the J higher level units, which follows from the assumption that the higher-level units are mutually independent observations. With $\boldsymbol{\eta}_{ij}^{(2)}$ and $\boldsymbol{\eta}_{j}^{(3)}$ being continuous latent variables at the lower and higher level of the analysis, $f(\mathbf{y}_j)$ is given by:

$$f(\mathbf{y}_j) = \int_{\boldsymbol{\eta}_j^{(3)}} \left[\prod_{i=1}^{n_j} f(\mathbf{y}_{ij} | \boldsymbol{\eta}_j^{(3)}) \right] f(\boldsymbol{\eta}_j^{(3)}) d\boldsymbol{\eta}_j^{(3)}, \tag{17}$$

where, as can be seen, the n_j observations within group j are assumed to be independent of one another given the group-level random coefficients $\boldsymbol{\eta}_j^{(3)}$. For each first-level unit, $f(\mathbf{y}_{ij}|\boldsymbol{\eta}_j^{(3)})$ is expressed by:

$$f(\mathbf{y}_{ij}|\boldsymbol{\eta}_j^{(3)}) = \int_{\boldsymbol{\eta}_{ij}^{(2)}} \left[\prod_{h=1}^{H} f(y_{hij} | \boldsymbol{\eta}_{ij}^{(2)}, \boldsymbol{\eta}_j^{(3)}) \right] f(\boldsymbol{\eta}_{ij}^{(2)} | \boldsymbol{\eta}_j^{(3)}) d\boldsymbol{\eta}_{ij}^{(2)}. \tag{18}$$

where $f(y_{hij}|\boldsymbol{\eta}_{ij}^{(2)}, \boldsymbol{\eta}_j^{(3)})$ is the distribution of the response variable h conditional on the lower and higher latent variables. The product over the H response variables follows from the local independence assumption.

In the MMFM we have a discrete latent variable at the higher level instead of continuous factors, which implies that the integration over $\boldsymbol{\eta}_j^{(3)}$ should be replaced by a summation over K classes. The relevant expression to construct the likelihood function are then:

$$\begin{aligned} f(\mathbf{y}_j) &= \sum_{k=1}^{K} \left[\prod_{i=1}^{n_j} f(\mathbf{y}_{ij} | \eta_{kj}^{(3)} = 1) \right] \pi_k \\ f(\mathbf{y}_{ij} | \eta_{kj}^{(3)} = 1) &= \int_{\boldsymbol{\eta}_{ij}^{(2)}} \left[\prod_{h=1}^{H} f(y_{hij} | \boldsymbol{\eta}_{ij}^{(2)}, \eta_{kj}^{(3)} = 1) \right] f(\boldsymbol{\eta}_{ij}^{(2)} | \eta_{kj}^{(3)} = 1) d\boldsymbol{\eta}_{ij}^{(2)}. \end{aligned}$$

In order to find the ML estimates, we need to solve the integrals involved in the computation of the likelihood function, as well as to maximize the marginal likelihood function. A closed form expression for the integrals is available only when both the indicators and the factors are assumed to be normally distributed. In all other situations, approximation methods should be used such as Laplace integration, numerical integration using (adaptive) quadrature, and Monte Carlo integration (Skrondal & Rabe-Hesketh, 2004). The most common algorithms for the maximization of likelihood function of models with latent variables are the Expectation-Maximization (EM) algorithm, Newton-Raphson (NR), Fisher scoring, and Quasi-Newton methods.

The three main software packages for latent variable modeling – GLLAMM (Rabe-Hesketh et al., 2004), MPlus (Muthén & Muthén, 1998-2007) and Latent GOLD (Vermunt & Magidson, 2005, 2008) – use slightly different combinations of the integration and maximization methods. GLLAMM uses numerical integration with either adapative or non-adaptive Gauss-Hermite quadrature and maximizes the marginal likelihood function using NR. In Mplus, the numerical integration is carried out using rectangular, Gauss-Hermite, or Monte Carlo integration, and the optimization is done using a combination of EM and a quasi-Newton method. Latent GOLD solves the integrals using Gauss-Hermite integration, and uses a combination of EM and NR to find the ML estimates.

After estimating the parameters of a MFM or MMFM, one will typically wish to obtain estimates of factor scores or latent class memberships. The most used factor scoring and class assignment methods make use of the posterior distribution of the latent variable(s) conditional on the observed data. It can be obtained using the well-known Bayes rule. At the higher level of the analysis, the conditional distribution of the latent variables given the observed data for group j is expressed by:

$$f(\boldsymbol{\eta}_j^{(3)}|\mathbf{y}_j) = \frac{f(\mathbf{y}_j, \boldsymbol{\eta}_j^{(3)})}{f(\mathbf{y}_j)} = \frac{f(\mathbf{y}_j|\boldsymbol{\eta}_j^{(3)})f(\boldsymbol{\eta}_j^{(3)})}{\int_{\boldsymbol{\eta}_j^{(3)}} f(\mathbf{y}_j|\boldsymbol{\eta}_j^{(3)})f(\boldsymbol{\eta}_j^{(3)})} \tag{19}$$

where $f(\mathbf{y}|\boldsymbol{\eta}_j^{(3)})$ and $f(\boldsymbol{\eta}_j^{(3)})$ are quantities defined by the estimated model.

The most widely used method for factor scoring is empirical Bayes (EB) prediction. It equals the mean of the posterior distribution of the latent variables defined in equation (19); that is:

$$\boldsymbol{\eta}_j^{(3)EB} = E(\boldsymbol{\eta}_j^{(3)}|\mathbf{y}_j) = \int_{\boldsymbol{\eta}_j^{(3)}} \boldsymbol{\eta}_j^{(3)} f(\boldsymbol{\eta}_j^{(3)}|\mathbf{y}_j). \tag{20}$$

With normally distributed latent variables, the empirical Bayes predictor is the best linear unbiased predictor BLUP (Skrondal & Rabe-Hesketh, 2004). In (multilevel) factor models with normally distributed indicators, the posterior density is multivariate normal and the posterior means can be obtained in closed form. For other response types, the posterior density gets closer to multivariate normal as the number of higher-level units increases (Skrondal & Rabe-Hesketh, 2004). As in the computation of the likelihood function, with non normal responses, the posterior distribution and means can not be expressed in closed form and numerical integration is thus required.

For the MMFM, which contains a discrete latent variable at the higher level, the posterior distribution is defined as follows:

$$P(\eta_{kj}^{(3)} = 1|\mathbf{y}_j) = \frac{f(\mathbf{y}_j, \eta_{kj}^{(3)} = 1)}{f(\mathbf{y}_j)} = \frac{f(\mathbf{y}_j|\eta_{kj}^{(3)} = 1)\,\pi_k}{\sum_{k=1}^{K} f(\mathbf{y}_j|\eta_{kj}^{(3)} = 1)\,\pi_k}. \tag{21}$$

The standard classification method in mixture models is the empirical Bayes modal (EBM) or posterior mode prediction; that is,

$$\eta_{kj}^{(3)EBM} = 1 \text{ if } \max_k P(\eta_{kj}^{(3)} = 1|\mathbf{y}_j). \tag{22}$$

3.2 Model evaluation

Model selection methods are needed to choose between MFMs and MMFMs with different structures; that is, to decide about number of factors, factor loading constraints, number of latent classes, etc. The common approach to compare nested models is by means of likelihood-ratio tests (Agresti, 2002). However, because certain regularity conditions do not hold, the underlying asymptotic theory does not apply when the null hypothesis of a test lies on the boundary of the parameter space. Two relevant examples in our context are tests comparing models with different numbers of factors and tests comparing models with different numbers of classes.

An alternative approach for comparing nested and non-nested models is via information criteria, which are basically penalized log-likelihood functions. These information criteria can be expressed as follows:

$$IC = -2 \ln L + C\, r \tag{23}$$

where $-2 \ln L$ is the minus twice the log-likelihood value that measures the fit of the model, r is the number of parameters and C is the penalty given for each additional parameter. The lower the value of an IC-like measure, the better the model. The various information criteria differ from one another in the choice of C. In the context of mixture modeling, a variety of textbooks and articles suggest using the Bayesian Information Criterion (BIC; Schwarz, 1978) for deciding about the number of latent classes (Nylund, Muthén & Asparouhov, 2007). The BIC is expressed by:

$$BIC = -2 \ln L + \ln(n)\, r, \tag{24}$$

thus uses the logarithm of the sample size ($\ln n$) as the penalty for each additional parameter. However, a complication factor in multilevel models is that the sample size could either be the number of individuals or the number of groups. But as shown in a recent simulation study by Lukočienė and Vermunt (2010), when comparing multilevel mixture models differing only at the between-level of analysis, as is the case of the application illustrated in the next paragraph, the number of groups is the better choice (see also Palardy & Vermunt, in press).

4 An application

In this section we present an application illustrating the MMFM as well as comparing it with the MFM. The analysis concerns a data set on students' satisfaction with the University of Florence collected one month before graduation (AlmaLaurea, 2005). The 2004 data set contains information on 1473 students attending 38 different study programs. The nine questionnaire items included in the analysis are student's satisfaction with 1) relationship with professors, 2) relationship with professors' assistants, 3) relationship with technical and administrative staff, 4) lecture rooms, 5) computers, 6) laboratories and facilities for the didactic activities, 7) libraries, 8) rooms for individual study and 9) global experience. The original items (except for items 5 and 8) contained four ordinal categories, but we collapsed the two lowest categories because the lowest category was seldom selected, which yielded: 1. Not satisfied or more not satisfied than satisfied, 2. More satisfied than not satisfied, 3. Definitively satisfied. Table 2 shows the item distributions.

[INSERT TABLE 2 ABOUT HERE]

The data set has a hierarchical structure (students are nested in study programs), which is why multilevel techniques should be used. Actually, the University of Florence is more interested in the comparison of programs than the comparison of students within programs, but nevertheless the within program factor model should be correctly specified. Below, we show that a the MFM yields interesting information on the dimensionality of between program variation in the students' satisfaction, whereas the MMFM yields a meaningful classification of programs. Combining the results of these two types of multilevel models yields a much better understanding of the phenomenon under study than each of the separate analysis.

Before applying the more complex factor models, we carried out exploratory analyses using a simple PCA and a set of univariate variance components models. The PCA with oblique rotation indicated that two components were needed: the first component concerns satisfaction with human aspects and global satisfaction (items 1, 2, 3, and 9) and the second component concerns physical aspects (items 4 to 8), where items 3 (relationship with technical and administrative staff) and 9 (global satisfaction) also have small loadings on the second factor.

Table 2 reports the value of the Intraclass Correlation Coefficient (ICC) obtained by running a two-level ordinal logistic regression model (with only a random intercept) for each of the nine items. Very strong program effects were found for the satisfaction with lecture rooms, computers, and library, moderate program effects for laboratories and relationship with technical staff, and weak program effects for the other four items. The fact that some of the ICCs are large indicates that a multilevel analysis makes very much sense. Moreover, the fact that the pattern of ICCs does not match the two-dimensional structure found with the PCA indicates that the latent structure at the program level is probably rather different from the latent structure at the student level.

Let us now turn to the analysis using the MFM. Because this concerns a model for ordinal indicators we used a proportional odds specification for the response model (see equation (3)). Moreover, we used the variant of the MFM with $\beta^{(3)}_{m_3 m_2} = 0$, which as explained earlier is the preferred specification when it cannot be assumed that the latent structures are the same at the student and program level. In equation (6), for each item h we constrained the cluster-level item-specific errors $e^{(3)}_{hj}$ to 0 to reduce the computational burden of the model (see also Grilli & Rampichini, 2007) and we fixed item intercepts μ_h to 0 for identification purposes. The items with the largest loading in the exploratory analysis (items 1 and 4) were used as the reference items (factor loading equal to 1) at the lower level and the item with largest ICC (item 5) as the reference item at the higher level.

First we estimated unrestricted factor models with 1, 2 and 3 factors at the student level, ignoring the multilevel structure[1]. Based on the BIC, we selected the model with 2

[1]The models were estimated using the Syntax module the Latent GOLD program (Vermunt & Magidson, 2008). We used Gauss-Hermite numerical integration with 10 quadrature nodes per dimensions. The

factors. Likelihood-ratio tests showed that various loadings could be set to 0, yielding a structure similar to the one expected based on the exploratory analysis using PCA. As a next step, we estimated variants of this restricted factor model with 1 and 2 higher-level factors. The BIC showed that only 1 factor was needed at the program level. Also for the higher-level factor various loadings could be equated to 0. The log-likelihood value of the final model equals -10195.14, the number of parameters is 37, and the BIC value is 20660.67 or 20525.33 depending on whether one uses a sample size equal to the number of individuals (1473) or groups (38). We also estimated the (2 factor) variance components model, for which we found larger BIC values showing that the variant of the MFM we used for our analysis fits better that the variance components factor model.

[INSERT TABLE 3 ABOUT HERE]

The parameters of the final model reported in Table 3 show that the two student-level factors represent the satisfaction with the Human Environment and the Physical Environment, respectively. Items 1, 2, 3, and 9 load on the first factor and items 3 to 9 on the second factor. As can be seen, satisfaction with the relationship with technical staff and global satisfaction load on both factors. Moreover, the Human Environment factor has a stronger effect on the students' global satisfaction that the Physical Environment factor. At the program level, we found only one factor which turns out to be related to the five Physical Environment items (items 4 to 8), to relationship with technical staff (item 3), and, although weakly, also to global satisfaction.

Rather than with a MFM, the same data set could also be analyzed with a MMFM, that is by using program-level latent classes instead of a factor. At the student level, we used the same factor structure as in the MFM; that is, a model with two correlated factors with some factor loadings constrained to 0. By using the constraint $\beta^{(3)}_{km_2} = 0$ in equation

number of starting sets was set to 50 with 50 EM iterations per set during the first stage.

(14) for each k and m_2, we assumed that the latent classes affect only the indicator means and thus not the lower-level factor means. As in the MFM, for each item h, we fixed the cluster-level item-specific errors $e_{hj}^{(3)}$ to 0 to reduce the computational burden and μ_h to 0 for identification purposes. Another identifying restriction on the parameters of equation (13) was $\sum_{k=1}^{K} \lambda_{kh}^{(3)} = 0$, which amounts to using effects coding for the classes. Table 4 shows BIC values for models with 1 to 7 latent classes. Based on the BIC using n equal to the number of groups, we selected the model with 5 classes.

[INSERT TABLES 4 AND 5 HERE]

[INSERT FIGURE 3 ABOUT HERE]

The estimates for the student-level part of the 5-class MMFM reported in Table 5 are almost identical to those obtained with the final MFM. At the program-level, we encountered 5 classes with class proportions equal to 0.084, 0.203, 0.427, 0.054, and 0.232. The estimates of $\lambda_{kh}^{(3)}$ parameters (see equation 13) are depicted in Figure 3. Because these are effect coded logit coefficients, positive values indicate that the latent class concerned is more satisfied than average and negative values that it is less satisfied than average. Programs belonging to Class 5 have the highest overall satisfaction since all its coefficients are positive and, moreover, largest for most of the items. On the contrary, in Class 1 all coefficients are negative, except the parameters for satisfaction with the individual spaces. Classes 2, 3 and 4 can be considered to be in between classes 1 and 5: while some of their coefficients are near to 0, others are either negative or positive showing (dis)satisfaction with specific issues. What can be observed is that the coefficients of Classes 2 and 3 have opposite signs, except for the items related to the library and individual spaces for which both classes show satisfaction. Class 4 is the most dissatisfied with the library and individual spaces. The larger variation of the $\lambda_{kh}^{(3)}$ parameters for the indicators related to Physical Environment (except for the one on individual spaces) confirms the result obtained with the MFM as well as with the exploratory analysis; i.e., programs differ most strongly on these items. The $\lambda_{kh}^{(3)}$ estimates also show how the MMFM results may complement the MFM results, especially because the former model is able to pick up specific differences between programs which are difficult to pick up with a factor analytic structure.

[INSERT FIGURE 4 ABOUT HERE]

As next step, we obtained empirical Bayes predictions for the program level latent factor using the MFM results and the empirical Bayes modal classifications using the MMFM results. Figure 4 depicts the 38 programs' class memberships on the x-axis and the programs' factor scores on the y-axis. A similar type of figure was used by Muthén (2001) to show the connection between the results of a standard FA with those of a standard mixture model. As can be seen, Classes 3 and 5 contain the programs with the higher factor scores and Classes 1, 2, and 4 the ones with the lower scores. However, there are also important differences between the programs that are captured by one approach but not by the other. For example, the programs "sc. economia e gestione aziendale" and "sc. economiche" (both from the Faculty of Economics), have very similar factor scores (0.44 and 0.51), which indicates similar overall satisfaction levels, but according to the MMFM results they belong to different classes: "sc. economia e gestione aziendale" is in Class 3 and "sc. economiche" is in Class 5, classes with similar global satisfaction, but with very different satisfaction with professors, lecture rooms, laboratories and computers. Thus, even if the overall satisfaction levels are similar, specific critical aspects of the two programs may differ. On the other hand, by looking only at the classification of the programs, one neglects possible differences within classes. For example, "lettere" and "scienze dei beni culturali" (both from the Faculty of Humanities) are both assigned to Class 2, but have rather different factor scores (-1.48 and -0.41), which shows that programs belonging to the same latent class may have rather different overall satisfaction levels.

5 Discussion

We proposed a factor analytic model for multilevel data sets containing a categorical latent variable at the higher level and thus yielding a clustering of higher-level units based on the parameters of the lower-level factor analytic model. We compared this new approach with the standard multilevel factor model, as well as showed how the two approaches may complement one another in empirical applications.

The proposed model can be extended in different manners. Whereas we used either a set of continuous latent variables or one discrete latent variable at each of the two levels, it is also possible to use combinations of continuous and discrete latent variables at both levels. This would yield a multilevel extension of the factor mixture model described among others by Lubke and Muthén (2005). A similar type of model was proposed by Palardy and Vermunt (in press) in the context of growth modeling. It is clear that such complex extensions should be handled with care because all kinds of practical problems may occur in their application. Note that existing latent variable modeling software allows estimating such more complex models.

Another important extensions of the models proposed in this article is the inclusion of lower-level and higher-level explanatory variables. For example, our application could be extended by including program-level explanatory variables explaining the class membership. Also this extension is readily implemented using existing latent variable modeling software.

One issue that clearly deserves additional attention is the issue of goodness- of-fit testing. The information criteria we used for model selection compare the overall fit of one model with that of other models. An alternative is to use fit measures quantifying specific goodness-of-fit aspects of a model. An example is a chi-squared statistic for the goodness-of-fit in the two-way table cross-tabulating two items, which is sometimes referred to as a bivariate residual statistic (Jöreskog & Moustaki, 2001; Vermunt & Magidson, 2005). Whereas such measures have been proposed for one-level factor and mixture models, they are not available yet for multilevel factor and mixture models.

References

Agresti, A. (2002). *Categorical data analysis*. Second edition. New York: Wiley.

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, *55*, 117-128.

AlmaLaurea (2005). *Profilo dei Laureati 2004*. Consorzio Interuniversitario. Bologna: AlmaLaurea.

Asparouhov, T., & Muthén, B. (2008). Multilevel mixture models. In Hancock, G. R. & Samuelsen, K. M. (Eds.), *Advances in latent variable mixture models*, 27-52. Charlotte, NC: Information Age Publishing, Inc.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis*. London: Arnold.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: J. WIley.

Bozdogan, H. (1987). Model selection and Akaikes information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*, 345-370.

Fox, J. P., & Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, *66*, 269-286.

Goldstein, H. (2003). *Multilevel statistical models*. Third edition. London: Arnold.

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

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61,* 215-231.

Grilli, L., & Rampichini, C. (2007). Multilevel factor models for ordinal variables. *Structural Equation Modeling*, *14*, 1-25.

Hox, J. (2002). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum.

Jöreskog, K., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, *36*, 347-387.

Lazarsfeld, P.F., & Henry, N.W. (1968). *Latent structure analysis.* Boston: Houghton Mill.

Longford, N., & Muthén, B. O. (1992). Factor analysis for clustered observations. *Psychometrika*, *57*, 581-597.

Lubke, G. H., & Muthén, B. O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*, 21-39.

Maij-de Meij, A.M., Kelderman, H., & van der Flier, H. (2008). Fitting a mixture item response theory model to personality questionnaire data: Characterizing latent classes and investigating possibilities for improving prediction. *Applied Psychological Measurement, 32*, 611-631.

McLachlan, G., & Peel, D. (2000). *Finite Mixture Models*. New York: John Wiley.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psycometrika*, *58*, 525-543.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*, 115-132.

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, *28*, 338-354.

Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (eds.), *New Developments and Techniques in Structural Equation Modeling* (pp. 1-33), Lawrence Erlbaum Associates.

Muthén, B. O. (2008). Latent variable hybrids: Overview of old and new models. In Hancock, G.R. & Samuelsen, K.M. (eds.) *Advances in latent variable mixture models* (pp. 1-24). Charlotte, NC: Information Age Publishing, Inc.

Muthén, B. O., & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In Fitzmaurice, G., Davidian, M., Verbeke, G. & Molenberghs, G. (eds.), *Longitudinal Data Analysis* (pp. 143-165). Boca Raton: Chapman & Hall/CRC Press.

Muthén, B. O., & Asparouhov, T. (2009). Multilevel regression mixture analysis. *Journal of the Royal Statistical Society, series A* , *172*, 639-657.

Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus User's Guide. Fourth Edition.*. Los Angeles, CA.

Muthén, B. O., & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. In D. R. Bock (Ed.), *Multilevel Analysis of Educational Data* (pp. 87-99), San Diego, CA: Academic Press.

Nylund, K. L., Muthén, B. O., & Asparouhov, T. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*, 535-569.

Palardy, G., & Vermunt, J. K. (in press). Multilevel growth mixture models for classifying group-level observations. *Journal of Educational and Behavioral Statistics.*

Rabe-Hesketh, S., & Skrondal, A. & Pickles, A. (2004). Generalized multilevel structural equation modelling. *Psychometrika*, *69*, 167-190.

Rost, J. (1990). Rasch model in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, *14*, 271-282

Searle S. R., Casella G., & McCullogh, C. E. (1992). *Variance Components.* New York: Wiley.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models.* Boca Raton, FL: Chapman & Hall/ CRC Press.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modelling.* London: Sage.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461-464.

Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology*, *33*, 213-239.

Vermunt, J. K. (2008a). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*, *17*, 33-51.

Vermunt, J. K. (2008b). Multilevel latent variable modeling: An application in education testing. *Austrian Journal of Statistics*, *37*, 285-299.

Vermunt, J. K., & Magidson, J. (2005). *Technical Guide for Latent GOLD 4.0: Basic and Advanced*. Belmont, Massachussetts: Statistical Innovations Inc.

Vermunt, J. K., & Magidson, J. (2008). *LG-Syntax User's Guide: Manual for Latent GOLD 4.5 Syntax Module*. Belmont, Massachussetts: Statistical Innovations Inc.

Vermunt, J. K., & van Dijk, L. (2001). A nonparametric random coefficients approach: the latent class regression model. *Multilevel Modelling Newsletter*, *13*, 6-13.

Yung, Y.F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, *62*, 297-330.

Table 1: Matrix of potential two-level models with underlying latent variables.

Lower level latent variables	Higher level latent variables	
	Discrete	Continuous
Discrete	I. Multilevel mixture latent class model	II. Multilevel latent class model
Continuous	III. Multilevel mixture factor model (MFM)	IV. Multilevel factor model (MMFM)

Table 2: Item frequency distributions (percentages). Univariate ordinal logit random intercept models, ICC. Students graduated at the University of Florence, year 2004.

Item	Level of satisfaction			# respondents	ICC
	1	2	3		
Rel. professors	16.7	65.6	17.8	1463	6.5
Rel. prof. assistants	20.7	58.3	21.1	1433	6.0
Rel. technical staff	43.4	43.0	13.6	1460	11.0
Lecture rooms	44.6	39.2	16.3	1451	23.3
Computers	14.8	66.0	19.3	1360	24.0
Laboratories	54.0	35.6	10.4	1143	24.1
Library	18.1	61.5	20.4	1351	10.4
Ind. spaces	14.6	55.3	30.1	1262	6.4
Global satisfaction	15.3	57.2	27.5	1468	6.3

Table 3: Two-level factor model: parameter estimates. Students graduated at the University of Florence, year 2004.

Item	Loadings			Thresholds	
	Within		Between		
	Human environment	Physical environment	Physical environment	α_1	α_2
Rel. professors	1			6.00	-6.00
Rel. prof. assistants	0.30			2.19	-2.23
Rel. techn. staff	0.14	0.28	0.25	0.49	-2.36
Lecture rooms		1	1.39	1.14	-2.69
Library		0.48	0.55	2.18	-1.56
Laborat.		0.86	0.81	0.14	-3.31
Computers		0.48	1	2.94	-1.59
Ind. spaces		0.34	0.36	2.21	-0.81
Global satisfaction	0.16	0.17	0.13	2.27	-1.25
Factor variance	49.98	6.11	0.82		
Factor correlation	0.48				

Table 4: Two-level mixture factor model: loglikelihood and BIC values. Students graduated at the University of Florence, year 2004.

Class	# Param.	Loglikelihood	BIC (# students)	BIC (# programs)
1	30	-10348.51	20915.88	20806.15
2	40	-10209.20	20710.20	20563.90
3	50	-10117.52	20599.79	20416.91
4	60	-10082.69	20603.08	20383.63
5	70	-10061.37	20633.39	20377.37
6	80	-10045.48	20674.56	20381.97
7	90	-10028.40	20713.36	20384.19

Table 5: Two-level mixture factor model: parameter estimates. Students graduated at the University of Florence, year 2004.

Item	Loadings Within: Human environment	Loadings Within: Physical environment	Thresholds α_1	Thresholds α_2
Rel. professors	1		6.05	-5.68
Rel. prof. assistants	0.34		2.21	-2.25
Rel. techn. staff	0.16	0.28	0.23	-2.71
Lecture rooms		1	0.23	-3.57
Library		0.53	1.82	-2.07
Laborat.		0.90	-0.16	-3.84
Computers		0.51	2.31	-2.25
Ind. spaces		0.37	1.97	-1.14
Global satisfaction	1.01	0.18	2.28	-1.29
Factor variance	42.23	5.78		
Factor correlation	0.47			

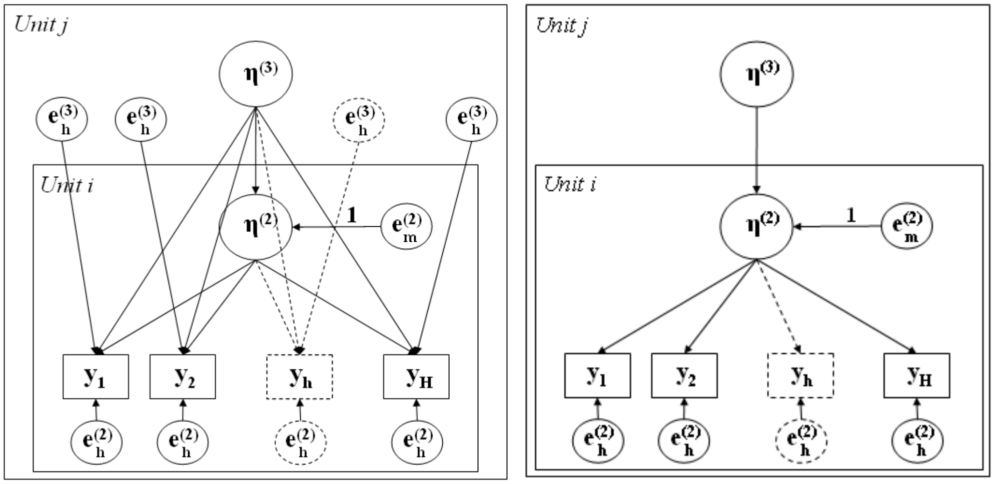

Figure 1: Two-level factor model (a) and variance component model (b)

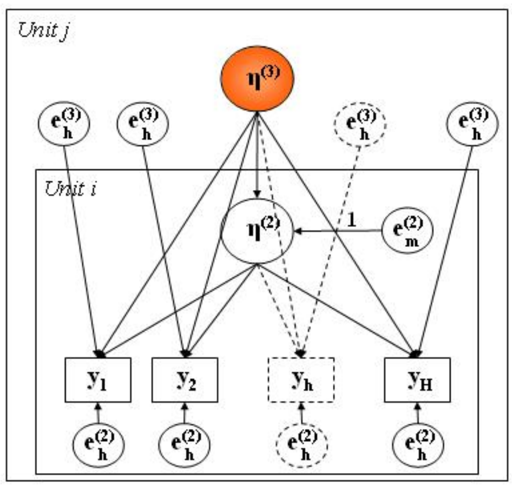

Figure 2: Two-level mixture factor model.

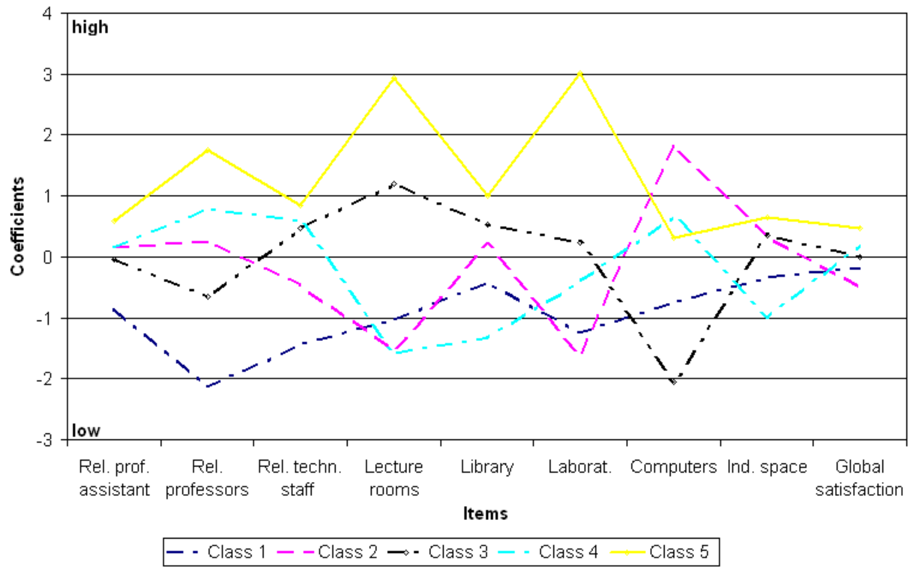

Figure 3: Two-level mixture factor model: latent classes features. Students graduated at the University of Florence, year 2004.

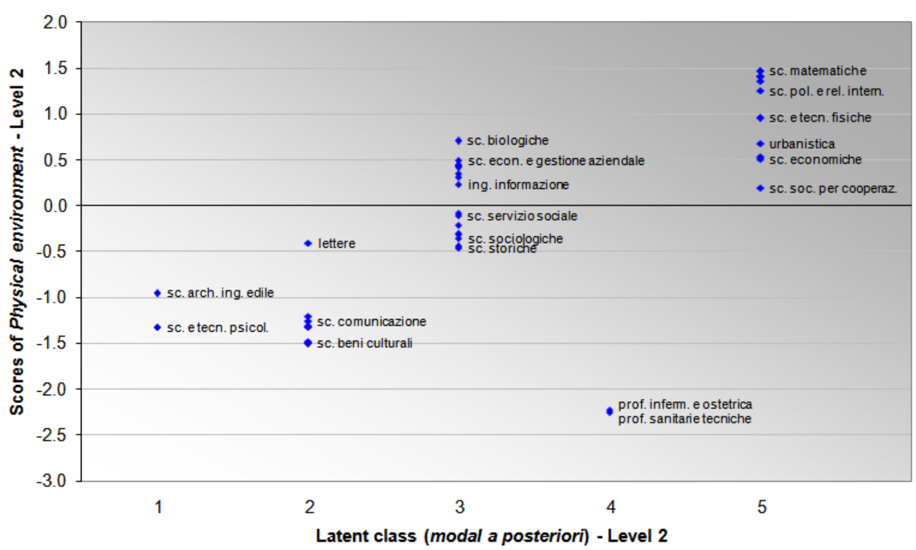

Figure 4: Two-level factor model and two-level mixture factor model: study programs factor scores and classes. Students graduated at the University of Florence, year 2004.