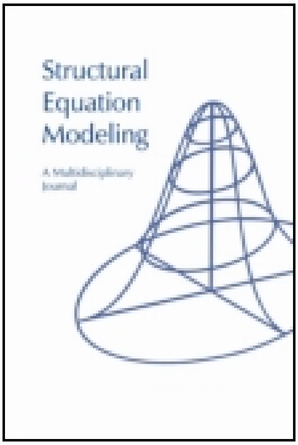

Structural Equation Modeling: A Multidisciplinary Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hsem20

Multilevel Factor Models for Ordinal Variables

Leonardo Grilli [a] & Carla Rampichini [a]
[a] Department of Statistics , University of Florence
Published online: 05 Dec 2007.

To cite this article: Leonardo Grilli & Carla Rampichini (2007) Multilevel Factor Models for Ordinal Variables, Structural Equation Modeling: A Multidisciplinary Journal, 14:1, 1-25

To link to this article: http://dx.doi.org/10.1080/10705510709336734

STRUCTURAL EQUATION MODELING, *14*(1), 1–25

Multilevel Factor Models for Ordinal Variables

Leonardo Grilli and Carla Rampichini
Department of Statistics, University of Florence

This article tackles several issues involved in specifying, fitting, and interpreting the results of multilevel factor models for ordinal variables. First, the problem of model specification and identification is addressed, outlining parameter interpretation. Special attention is devoted to the consequences on interpretation stemming from the usual choice of not decomposing the specificities into hierarchical components. Then a general strategy of analysis is outlined, highlighting the role of the exploratory steps. The theoretical topics are illustrated through an application to graduates' job satisfaction, where estimation is based on maximum likelihood via an Expectation-Maximization algorithm with adaptive quadrature.

The classical factor model is a popular and effective tool of analysis in the social sciences. In its standard formulation (e.g., Anderson, 2003), it concerns a set of continuous variables measured on a set of independent units. However such features may be inadequate in many cases, for example, when the response variables are measured on ordinal scales (e.g., Likert scales) or the statistical units are nested in multilevel structures.

In principle, the specification of a factor model when the variables are ordinal (Jöreskog & Moustaki, 2001; Moustaki, Jöreskog, & Mavridis, 2004) does not entail relevant theoretical problems, but estimation is more computationally demanding, so a number of solutions have been proposed, chiefly (a) transform the ordinal scale in a continuous scale by assigning a score to each grade, and use methods for continuous data (this procedure works well when there are enough categories, and the frequency distribution is unimodal with an internal mode; Muthén & Kaplan, 1985); (b) estimate the polychoric correlations among the ordinal variables and use these correlations as the input in algorithms for continuous data (this approach

Correspondence should be addressed to Leonardo Grilli, Dipartimento di Statistica "G. Parenti," Viale Morgagni, 59, 50134 Firenze, Italy. E-mail: grilli@ds.unifi.it

gives consistent point estimates, but it does not rely on a complete statistical model for the observed data, so it is not fully efficient). This second method, usually called weighted least aquares (Muthén, 1984; Muthén & Satorra, 1995), is available in many packages, including LISREL (Jöreskog, 2005; Jöreskog & Sörbom, 2005) and M*plus* (Muthén & Muthén, 2006).

Recent developments in computational statistics have greatly enhanced the feasibility of a maximum likelihood (ML) or Bayesian analysis based on the proper model. In particular, the software M*plus* and the Stata command `gllamm` (Rabe-Hesketh, Skrondal, & Pickles, 2004b) perform full information ML via numerical integration.

In many fields the statistical units are quite often nested in hierarchical or multilevel structures. Even if multilevel models are now well developed (Goldstein, 2003; Snijders & Bosker, 1999), the subclass of multilevel factor models has received relatively little attention, especially in applied work. For the case of continuous variables, two classical references are Goldstein and McDonald (1988) and Longford and Muthén (1992), whereas the case of binary variables has been treated only recently by Ansari and Jedidi (2000) and Goldstein and Browne (2005).

This article focuses on multilevel factor models for ordinal variables, a case where the problems associated with a proper treatment of ordinal variables add to the difficulties of a multilevel analysis. From a theoretical standpoint such a model is a member of some broad frameworks, such as the nonlinear mixed model framework for item response theory (IRT) analysis of Rijmen, Tuerlinckx, De Boerck, and Kuppens (2003), and the class of generalized linear latent and mixed models (GLLAMM) of Rabe-Hesketh, Skrondal, and Pickles (2004a) and Skrondal and Rabe-Hesketh (2004). However, the multilevel factor model for ordinal variables raises several identification and interpretation issues, as well as computational problems, so its implementation is not straightforward and, in fact, its use in applied work is rare.

This article discusses how to specify and fit multilevel factor models for ordinal variables, illustrating the theory by means of an application on the job satisfaction of the 1998 graduates of the University of Florence. The graduates responded to a series of items on job satisfaction using a 5-point ordinal scale, so the model should treat the responses as ordinal variables. The main aim of the analysis is to describe and summarize the aspects of job satisfaction measured by the considered items, separately for the graduate and degree program levels, to shed light on the effectiveness of the degree programs. Therefore, the hierarchical nature of the phenomenon, with graduates nested in degree programs, has a primary role and the use of a multilevel factor model is essential, as it allows definition of separate factor structures at the two levels.

All the computations needed in the application are performed using standard software. In particular, the model is fitted with the package M*plus,* which performs ML via an EM algorithm with adaptive Gaussian quadrature.

This article considers the case where all the response variables are ordinal. The case of binary variables is simply a special instance, and the extension to mixed continuous-ordinal-binary responses is straightforward.

The structure of the article is as follows. In the second section the model is defined, showing the likelihood and outlining the problem of identification; then the interpretation of model parameters is discussed, stressing the implications of not decomposing the specificities in the between and within components. The third section presents a general strategy of analysis with several exploratory steps, which is made necessary by the computational effort usually required to fit the model. In the fourth section an application to the analysis of job satisfaction of the 1998 graduates of the University of Florence illustrates the steps of analysis and the interpretation of the results. The final section concludes with some remarks.

THE TWO-LEVEL FACTOR MODEL FOR ORDINAL VARIABLES

Description of the Model

The analysis of ordinal data requires choosing a set of probabilities to be modeled and a suitable link function (Agresti, 2002). The cumulative probabilities are chosen here because this allows representation of each observed ordinal variable in terms of a continuous latent response endowed with a set of thresholds, a representation that helps the presentation of the model and the interpretation of the results. Note, however, that the latent response approach is only a convenient way to represent an ordinal variable; it does not require that the data have been generated by categorizing latent response variables.

The model can be formulated with any link function. In the following, the probit link is adopted because the latent response is assumed to be Gaussian, a quite natural choice to create a connection with the classical factor model. However, the use of other links entails few modifications and usually leads to negligible differences in the results.

Let Y_{hij} be the hth observed ordinal variable (item; $h = 1, 2, \dots , H$) for the ith subject ($i = 1, 2, \dots , n_j$) of the jth cluster ($j = 1, 2, \dots J$). Even if the items could be seen as lowest level units, here the hierarchical levels are numbered starting from the subjects, so the subjects are referred to as Level 1 or "within" units, whereas the clusters are referred to as Level 2 or "between" units. Attention is limited to two hierarchical levels, because the extension to more than two levels is conceptually straightforward. The model allows for item nonresponse; that is, for subject i of cluster j, Y_{hij} may be missing for some h. In the application, the clusters are the degree programs, the subjects are the graduates, and the ordinal variables are the ratings on five items of the questionnaire (i.e., $H = 5$).

A two-level factor model for ordinal variables can be set up by defining two components, namely: (a) a threshold model that relates a set of continuous latent variables $\widetilde{Y}_{hij}$ to the observed ordinal counterparts Y_{hij}, and (b) a two-level factor model for the set of continuous latent variables $\widetilde{Y}_{hij}$.

As for the threshold model, assume that each of the observed responses Y_{hij}, which takes values in $\{1, 2, \ldots, C_h\}$, is generated by a latent continuous variable $\widetilde{Y}_{hij}$ through the following relation:

$$\left\{Y_{hij} = c_h\right\} \Leftrightarrow \left\{\gamma_{c_h-1,h} < \tilde{Y}_{hij} \leq \gamma_{c_h,h}\right\}, \tag{1}$$

where the thresholds satisfy $-\infty = \gamma_{0,h} \leq \gamma_{1,h} \leq \ldots \leq \gamma_{C_h-1,h} \leq \gamma_{C_h,h} = +\infty$.

The factor model can now be defined on the set of latent variables. Ignoring for the moment the hierarchical structure, the standard factor model can be written as

$$\tilde{Y}_{hij} = \mu_h + \sum_{m=1}^{M} \lambda_{mh} u_{mij} + e_{hij}, \qquad h = 1, \cdots, H \tag{2}$$

where u_{mij} are the common factors, and, for each h, μ_h is the item mean, λ_{mh} are the factor loadings and e_{hij} are the uncorrelated item-specific errors.

A two-level extension of the factor model can be obtained in two different ways (see, e.g., Muthén, 1994). The simplest way is to decompose the factors and the item-specific errors in two components, one for each hierarchical level:

$$\begin{aligned} u_{mij} &= u_{mj}^{(2)} + u_{mij}^{(1)}, \qquad && m = 1, \cdots, M, \\ e_{hij} &= e_{hj}^{(2)} + e_{hij}^{(1)}, \qquad && h = 1, \cdots, H. \end{aligned}$$

Here we use the superscript (l) to denote random variables defined at level l, $l = 1$, 2, along with their parameters and loadings. Therefore, Model 2 becomes

$$\begin{aligned} \tilde{Y}_{hij} &= \mu_h + \sum_{m=1}^{M} \lambda_{mh} \left(u_{mj}^{(2)} + u_{mij}^{(1)}\right) + \left(e_{hj}^{(2)} + e_{hij}^{(1)}\right) \\ &= \mu_h + \left[\sum_{m=1}^{M} \lambda_{mh} u_{mj}^{(2)} + e_{hj}^{(2)}\right] + \left[\sum_{m=1}^{M} \lambda_{mh} u_{mij}^{(1)} + e_{hij}^{(1)}\right]. \end{aligned} \tag{3}$$

This formulation is useful if one assumes the existence of certain factors and wishes to study how they vary between and within the clusters. However, in applied work it is common to find completely different factor structures at the two hierarchical levels, so a more general formulation is this (Goldstein & McDonald, 1988; Longford & Muthén, 1992):

$$\tilde{Y}_{hij} = \mu_h + \left[\sum_{m=1}^{M_2} \lambda_{mh}^{(2)} u_{mj}^{(2)} + e_{hj}^{(2)}\right] + \left[\sum_{m=1}^{M_1} \lambda_{mh}^{(1)} u_{mij}^{(1)} + e_{hij}^{(1)}\right]. \tag{4}$$

In this model the cluster level has M_2 factors with corresponding loadings $\lambda_{mh}^{(2)}$, and the subject level has M_1 factors with corresponding loadings $\lambda_{mh}^{(1)}$. Note that, even when $M_2 = M_1$, the factor loadings in general are different, so the factors may have different interpretations. Obviously, Model 3 is a special case of Model 4 with $M_2 = M_1$ and $\lambda_{mh}^{(2)} = \lambda_{mh}^{(1)}$.

Now it is convenient to express the general two-level model (Equation 4) for the latent responses in matrix notation:

$$\tilde{\mathbf{Y}}_{ij} = \boldsymbol{\mu} + \left[\boldsymbol{\Lambda}^{(2)}\mathbf{u}_j^{(2)} + \mathbf{e}_j^{(2)}\right] + \left[\boldsymbol{\Lambda}^{(1)}\mathbf{u}_{ij}^{(1)} + \mathbf{e}_{ij}^{(1)}\right], \tag{5}$$

where $\tilde{\mathbf{Y}}_{ij} = (\tilde{Y}_{1ij}, \cdots, \tilde{Y}_{Hij})'$, $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_H)'$, $\mathbf{e}_j^{(2)} = (e_{1j}^{(2)}, \cdots, e_{Hj}^{(2)})'$, $\mathbf{u}_j^{(2)} =$, $(u_{1j}^{(2)}, \cdots, u_{M_2 j}^{(2)})'$ $\mathbf{e}_{ij}^{(1)} = (e_{1ij}^{(1)}, \cdots, e_{Hij}^{(1)})'$, $\mathbf{u}_{ij}^{(1)} = (u_{1ij}^{(1)}, \cdots, u_{M_1 ij}^{(1)})'$, while $\boldsymbol{\Lambda}^{(2)}$ is a matrix whose hth row is $(\lambda_{1h}^{(2)}, \cdots, \lambda_{M_2 h}^{(2)})$ and $\boldsymbol{\Lambda}^{(1)}$ is a matrix whose hth row is $(\lambda_{1h}^{(1)}, \cdots, \lambda_{M_1 h}^{(1)})$.

The standard assumptions on the item-specific errors of Model 5 are

$$\mathbf{e}_j^{(2)} \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Psi}^{(2)}), \qquad \boldsymbol{\Psi}^{(2)} = diag\{(\psi_h^{(2)})^2\},$$

and

$$\mathbf{e}_{ij}^{(1)} \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Psi}^{(1)}), \qquad \boldsymbol{\Psi}^{(1)} = diag\{(\psi_h^{(1)})^2\},$$

and for the factors it is assumed that

$$\mathbf{u}_j^{(2)} \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}^{(2)}),$$

and

$$\mathbf{u}_{ij}^{(1)} \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}^{(1)}),$$

where the covariance matrices $\boldsymbol{\Sigma}^{(2)}$ and $\boldsymbol{\Sigma}^{(1)}$ are, in principle, unconstrained, with diagonal elements representing factor variances $(\sigma_m^{(l)})^2$, $l = 1, 2$, $m = 1, \ldots, M_l$. Moreover, all the errors and factors are assumed to be mutually independent, except for the factors at the same level, so Model 5 is equivalent to the following variance decomposition

$$Var(\tilde{\mathbf{Y}}_{ij}) = \left[\boldsymbol{\Lambda}^{(2)}\boldsymbol{\Sigma}^{(2)}\boldsymbol{\Lambda}^{(2)'} + \boldsymbol{\Psi}^{(2)}\right] + \left[\boldsymbol{\Lambda}^{(1)}\boldsymbol{\Sigma}^{(1)}\boldsymbol{\Lambda}^{(1)'} + \boldsymbol{\Psi}^{(1)}\right]. \tag{6}$$

This amounts to a couple of factor models, one for the between covariance matrix and the other for the within covariance matrix (Muthén, 1994).

Finally note that the model can be extended by adding a regression component $\mathbf{x}'_{hij}\boldsymbol{\beta}$ in Equation 4. Although it does not alter the essence of the article, the introduction of a regression component gives rise to a wide range of models, depending on the nature of the covariates. The covariates can be of various types (Rijmen et al., 2003): item covariates, unit covariates, and item by unit covariates, where the unit covariates can be further distinguished into subject-level and cluster-level covariates.

Model Likelihood and Identification

The full likelihood for the two-level factor model (Equation 4) can be derived in the following steps. Denoting with $\boldsymbol{\theta}$ the set of estimable parameters, the conditional likelihood for subject i of cluster j is

$$L_{ij}(\boldsymbol{\theta}|\mathbf{u}_j^{(2)},\mathbf{e}_j^{(2)}) = E_{\mathbf{u}^{(1)}}\left[\prod_{\mathbf{c}\in\mathcal{C}}\left(P\left(\bigcap_{h=1}^{H}\{Y_{hij}=c_h\}\,|\,\mathbf{u}_{ij}^{(1)},\mathbf{u}_j^{(2)},\mathbf{e}_j^{(2)}\right)\right)^{d_{ij\mathbf{c}}}\right], \tag{7}$$

where C is the set of all admissible values of the vector $\mathbf{c} = (c_1, \ldots, c_h)'$ and $d_{ij\mathbf{c}}$ is the indicator of the observed response pattern $\left\{\bigcap_{h=1}^{H}\{Y_{hij} = y_{hij}\}\right\}$.

The overall marginal likelihood is then

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{J} E_{\mathbf{u}^{(2)},\mathbf{e}^{(2)}}\left[\prod_{i=1}^{n_j} L_{ij}(\boldsymbol{\theta}|\mathbf{u}_j^{(2)},\mathbf{e}_j^{(2)})\right]. \tag{8}$$

The probabilities that appear in Equation 7, given the relation (Equation 1) between the observed and latent responses and the assumptions on the latent model (Equation 4), can be written as

$$\begin{aligned}
&P\left(\bigcap_{h=1}^{H}\{Y_{hij}=c_h\}\,|\,\mathbf{u}_{ij}^{(1)},\mathbf{u}_j^{(2)},\mathbf{e}_j^{(2)}\right)\\
&= P\left(\bigcap_{h=1}^{H}\{\gamma_{c_h-1,h} < \tilde{Y}_{hij} \le \gamma_{c_h,h}\}\,|\,\mathbf{u}_{ij}^{(1)},\mathbf{u}_j^{(2)},\mathbf{e}_j^{(2)}\right)\\
&= \prod_{h=1}^{H} P\left(\gamma_{c_h-1,h} - \mu_h - \zeta_{hij} < e_{hij}^{(1)} \le \gamma_{c_h,h} - \mu_h - \zeta_{hij}\;|\,\mathbf{u}_{ij}^{(1)},\mathbf{u}_j^{(2)},\mathbf{e}_j^{(2)}\right)\\
&= \prod_{h=1}^{H}\left[\Phi\left(\frac{\gamma_{c_h,h}-\mu_h}{\psi_h^{(1)}} - \frac{\zeta_{hij}}{\psi_h^{(1)}}\right) - \Phi\left(\frac{\gamma_{c_h-1,h}-\mu_h}{\psi_h^{(1)}} - \frac{\zeta_{hij}}{\psi_h^{(1)}}\right)\right],
\end{aligned} \tag{9}$$

where $\zeta_{hij} = \sum_{m=1}^{M_2} \lambda_{mh}^{(2)} u_{mj}^{(2)} + e_{hj}^{(2)} + \sum_{m=1}^{M_1} \lambda_{mh}^{(1)} u_{mij}^{(1)}$ and $\Phi(\cdot)$ is the standard Gaussian distribution function. Note that $\Phi(\cdot)$ is the inverse of the link function and stems from the normality assumption on the subject-level item-specific errors $e_{hij}^{(1)}$. Other distributional assumptions lead to different link functions (e.g., the logistic distribution leads to the logit link).

In light of Equations 7 and 9, the likelihood (Equation 8) is equivalent to the likelihood of a three-level model, where the items are the first-level units, the subjects are the second-level units, and the clusters are the third level-units. This correspondence is useful for estimation purposes.

To discuss identification issues, note from Equation 9 that the model likelihood is based on the quantities

$$\begin{aligned} & \Phi\left(\frac{\gamma_{c_h,h} - \mu_h}{\psi_h^{(1)}} - \frac{\zeta_{hij}}{\psi_h^{(1)}}\right) \\ = \; & \Phi\left(\frac{\gamma_{c_h,h} - \mu_h}{\psi_h^{(1)}} - \sum_{m=1}^{M_2} \frac{\lambda_{mh}^{(2)} u_{mj}^{(2)}}{\psi_h^{(1)}} - \frac{e_{hj}^{(2)}}{\psi_h^{(1)}} - \sum_{m=1}^{M_1} \frac{\lambda_{mh}^{(1)} u_{mij}^{(1)}}{\psi_h^{(1)}}\right) \\ = \; & \Phi\left(\frac{\gamma_{c_h,h} - \mu_h}{\psi_h^{(1)}} - \sum_{m=1}^{M_2} \frac{\lambda_{mh}^{(2)} \sigma_m^{(2)}}{\psi_h^{(1)}} u_{mj}^{(2)*} - \frac{\psi_h^{(2)}}{\psi_h^{(1)}} e_{hj}^{(2)*} - \sum_{m=1}^{M_1} \frac{\lambda_{mh}^{(1)} \sigma_m^{(1)}}{\psi_h^{(1)}} u_{mij}^{(1)*}\right) \end{aligned} \tag{10}$$

for $h = 1, \dots H$ and $c_h = 1, \dots, C_h - 1$, where the asterisk denotes standardized variables. Assuming for simplicity that all the factors are uncorrelated, from Equation 10 the estimable quantities are:

- $(\gamma_{c_h,h} - \mu_h)/\psi_h^{(1)}$ in number of $\sum_{h=1}^{H} (C_h - 1)$
- $(\lambda_{mh}^{(2)} \sigma_m^{(2)})/\psi_h^{(1)}$ in number of HM_2.
- $(\psi_h^{(2)})/\psi_h^{(1)}$ in number of H.
- $(\lambda_{mh}^{(1)} \sigma_m^{(1)})/\psi_h^{(1)}$ in number of HM_1.

Note that all the estimable quantities are expressed in terms of the item-specific subject-level standard deviations $\psi_h^{(1)}$, but this is not a problem for interpretation. See the following subsection "Interpretation of Model Parameters."

The constraints needed for model identification are of two kinds: Some are needed for the identification of the distribution of the latent responses determining the ordinal measures (Equation 1), and others are needed for the identification of the covariance matrix decomposition (Equation 6) associated with the factor model.

The identification of the distribution of the latent responses $\tilde{Y}_{hij}$ yielding the observed responses Y_{hij} does not depend on the hierarchical nature of the model, so the considerations are the same as for single-level factor models for ordinal variables. First, recall that in a univariate ordinal model the mean and the standard deviation of the latent response are not identifiable, so it is necessary to use two constraints; for example $\mu_h = 0$ and $\psi_h^{(1)} = 1$. In a multivariate ordinal model (Grilli & Rampichini, 2003), a possibility is to impose constraints on the mean and standard deviation of each item and freely estimate all the thresholds of all the items. In such a case, the threshold model uses all the available degrees of freedom, so the factorial part is not threatened by potentially invalid restrictions on the threshold part. A useful feature of unconstrained thresholds is that all the specificities are equal, so the factor loadings of two items are directly comparable.

However, the number of freely estimable thresholds, $\sum_{h=1}^{H}(C_h - 1)$, is usually quite large, so the parsimony principle suggests to look for a constrained structure entailing a negligible loss of fit. In general, structuring the thresholds in a sensible way is not straightforward, but when all the items have the same categories ($C_h = C \geq 3$) an appealing option (later called *equal latent thresholds*) is to assume that at every cutpoint c the latent thresholds $\gamma_{c,h}$ are equal; that is, $\gamma_{c,h'} = \gamma_{c,h''} = \gamma_c$ for each h', h'' and leave μ_h and $\psi_h^{(1)}$ free (except for a reference item): In this way the actual thresholds are $\tau_{c,h} = (\gamma_c - \mu_h)/\psi_h^{(1)}$, as it is clear from the likelihood contribution (Equation 9). Obviously, this kind of restriction makes the threshold structure easily interpretable, although it requires some care in the interpretation of the loadings, as each item has its own scale $\psi_h^{(1)}$.

As for the identification of the two-level factor model on the continuous latent variables $\tilde{Y}_{hij}$, the variance–covariance decomposition (Equation 6) entails M_2^2 and M_1^2 indeterminacies in $\mathbf{\Lambda}^{(2)}\mathbf{\Sigma}^{(2)}\mathbf{\Lambda}^{(2)\prime} + \mathbf{\Psi}^{(2)}$ and $\mathbf{\Lambda}^{(1)}\mathbf{\Sigma}^{(1)}\mathbf{\Lambda}^{(1)\prime} + \mathbf{\Psi}^{(1)}$, respectively. In exploratory factor analysis it is customary to assume uncorrelated (orthogonal) factors with unit variance, putting the remaining constraints on the factor loadings in various forms (see, e.g., Anderson, 2003). However, in confirmatory factor analysis it may be useful to relax either or both the assumptions on the factor covariance matrix (i.e., unit variance and uncorrelatedness). Relaxing the unit variance assumption causes a scale indeterminacy that can be solved by fixing to one a loading for each factor, and the uncorrelatedness is usually compensated for by an adequate number of zeroes in the matrix of loadings. The main advantage of an unconstrained factor covariance matrix is that the loadings are invariant with regard to certain changes, as in the cases of factor-based unit selection and comparisons among populations (Anderson, 2003). However, it is clear from the following that correlated factors complicate the interpretation of the results, whereas unconstrained variances are harmless in this regard.

The analytical approach to identification just sketched can be formalized, in the structural equation framework, through the definition of identification mappings between the structural parameters and identified reduced form parameters, as in Skrondal and Rabe-Hesketh (2004). Finally, local identification can be empirically checked in the estimation phase by inspection of the rank of the ML information matrix (e.g., computing the condition number), as nonsingularity of the information matrix is a sufficient condition—although not necessary in the case of nonlinear models—for local identification (Skrondal & Rabe-Hesketh, 2004).

Interpretation of Model Parameters

The formerly outlined two-level factor model for ordinal variables is based on two components that can be interpreted separately: (a) a threshold model that relates the continuous latent responses $\tilde{Y}_{hij}$ to the observed ordinal counterparts Y_{hij}, and (b) a two-level factor model for the continuous latent responses $\tilde{Y}_{hij}$. The following discussion focuses on some issues concerning the second component, which conveys the most important information.

Although the interpretation of the two-level factor model relies on the classical ideas of factor analysis, some clarification may be useful. Note that the following formulas are based on the uncorrelatedness of the factors at both levels, whereas the factor variances may be fixed or free (in any case it is assumed that the model is identified through adequate constraints on the loadings). It should be noted that the factor variances are not directly interpretable, even when left free, as they simply represent contributions with respect to the arbitrary item that has the loading fixed to one: In general, the only interpretable quantity is the variance contribution expressed by the product $\left(\lambda_{mh}^{(1)}\sigma_m^{(1)}\right)^2$ or $\left(\lambda_{mh}^{(2)}\sigma_m^{(2)}\right)^2$.

From Equation 6, the total variance of the hth item is decomposed in

$$Var_T\left(\tilde{Y}_{hij}\right) = Var_B\left(\tilde{Y}_{hij}\right) + Var_W\left(\tilde{Y}_{hij}\right), \tag{11}$$

where

$$\begin{array}{rcccc} Var_B\left(\tilde{Y}_{hij}\right) & = & \displaystyle\sum_{m=1}^{M_2}\left(\lambda_{mh}^{(2)}\sigma_m^{(2)}\right)^2 & + & \left(\psi_h^{(2)}\right)^2 \\ & & \text{Cluster-level communality} & + & \text{Cluster-level specificity,} \end{array}$$

and

$$\begin{array}{rcccc} Var_W\left(\tilde{Y}_{hij}\right) & = & \displaystyle\sum_{m=1}^{M_1}\left(\lambda_{mh}^{(1)}\sigma_m^{(1)}\right)^2 & + & \left(\psi_h^{(1)}\right)^2 \\ & & \text{Subject-level communality} & + & \text{Subject-level specificity.} \end{array}$$

The ratio

$$ICC_h = \frac{Var_B\left(\tilde{Y}_{hij}\right)}{Var_T\left(\tilde{Y}_{hij}\right)} \tag{12}$$

is the so-called intraclass correlation coefficient (ICC), which represents the proportion of variance explained by the clusters.

In most applications, to save computational resources, the cluster-level item-specific errors $e_{hj}^{(2)}$ in Equation 4 are omitted. In such a case, the variance of the remaining subject-level item-specific errors $e_{hij}^{(1)}$ represents the total specificity, and the factor structure is unaffected (even if the estimates may change substantially if the factor structure is poorly specified). A consequence of omitting the errors $e_{hj}^{(2)}$ is that the variance decomposition (Equation 11) is not feasible, so it is important to understand the role of such decomposition for interpretation.

First, consider the case where the specificities are not disentangled. In general, the interpretation of the factor structures at the two levels does not depend on the decomposition of the specificities and the (*relative*) *communalities* can be computed as well. As in standard factor models, the *communality* is the proportion of the variance of a given response explained by the factors. As usual with ordinal items, the communalities are referred to as the *latent* responses. For example, the total communality of the *h*th item is

$$\frac{\sum_{m=1}^{M_1}\left(\lambda_{mh}^{(1)}\sigma_m^{(1)}\right)^2+\sum_{m=1}^{M_2}\left(\lambda_{mh}^{(2)}\sigma_m^{(2)}\right)^2}{Var_T\left(\tilde{Y}_{hij}\right)}, \tag{13}$$

and the communality of the *h*th item due to the *m*th subject-level factor is

$$\frac{\left(\lambda_{mh}^{(1)}\sigma_m^{(1)}\right)^2}{Var_T\left(\tilde{Y}_{hij}\right)}. \tag{14}$$

Moreover, the decomposition of the specificities is not required for the correlation between two latent responses of the same subject, $\tilde{Y}_{h'ij}$ and $\tilde{Y}_{h''ij}$:

$$\frac{\sum_{m=1}^{M_1}\lambda_{mh'}^{(1)}\lambda_{mh''}^{(1)}\left(\sigma_m^{(1)}\right)^2+\sum_{m=1}^{M_2}\lambda_{mh'}^{(2)}\lambda_{mh''}^{(2)}\left(\sigma_m^{(2)}\right)^2}{\sqrt{Var_T\left(\tilde{Y}_{h'ij}\right)\cdot Var_T\left(\tilde{Y}_{h''ij}\right)}}.$$

However, there are other interesting quantities that can be computed only if the specificities are disentangled, such as the ICC$_h$ (Equation 12) and the communalities at a given level. For example, the total communality at subject level of the *h*th item is

$$\frac{\sum_{m=1}^{M_1}\left(\lambda_{mh}^{(1)}\sigma_m^{(1)}\right)^2}{Var_W\left(\tilde{Y}_{hij}\right)},$$

and the communality at subject level of the *h*th item due to the *m*th subject-level factor is

$$\frac{\left(\lambda_{mh}^{(1)}\sigma_m^{(1)}\right)^2}{Var_W\left(\tilde{Y}_{hij}\right)}.$$

Moreover, for a given item, the correlation between two distinct subjects belonging to the same cluster is just the ICC_h (Equation 12), so it is computable only if the specificities are decomposed.

Finally, note from Equation 10 that all the estimable quantities are scaled by $\psi_h^{(1)}$. If the cluster-level item-specific errors $e_{hj}^{(2)}$ are omitted, each scale factor $\psi_h^{(1)}$ represents the square root of the item total specificity, leading to smaller estimable quantities. Nevertheless, the communalities are unaffected by the item scale, as they are ratios of parameters within the same item.

PHASES OF THE ANALYSIS

The accomplishment of careful exploratory analyses is extremely important to achieve a suitable model specification, helping to avoid some of the traps that inevitably characterize the development of a complex model. Moreover, fitting the two-level factor model for ordinal variables, outlined in the previous section, is computationally intensive. In fact, the marginal likelihood (Equation 8) involves multiple integrals with respect to Gaussian densities that cannot be solved analytically. Several estimation methods have been proposed, such as ML with adaptive Gaussian quadrature (Rabe-Hesketh, Skrondal, & Pickles, 2005) and Bayesian Markov Chain Monte Carlo (MCMC) algorithms (Ansari & Jedidi, 2000; Goldstein & Browne, 2002, 2005); but other methods can be successfully applied, as discussed in the final section. The computational burden is heavy, so it is crucial to base model selection on suitable exploratory analyses, thus limiting the number of fitted models and supplying the algorithms with good starting values.

For the analysis we suggest adapting to the ordinal case Muthén's (1994) strategy for continuous items.

1. Univariate two-level models. As a first step, it is advisable to fit a set of univariate ordinal random intercept models, one for each item, with the following specification in terms of latent responses:

$$\tilde{Y}_{hij} = \mu_h + e_{hj}^{(2)} + e_{hij}^{(1)}, \qquad h = 1, \cdots, H \quad , \tag{15}$$

where $e_{hj}^{(2)}$ are cluster-level errors with standard deviation $\psi_h^{(2)}$ and $e_{hij}^{(1)}$ are subject-level errors with standard deviation $\psi_h^{(1)}$, implying $Var(\tilde{Y}_{hij}) =$

$\left(\psi_h^{(2)}\right)^2 + \left(\psi_h^{(1)}\right)^2$. To overcome the usual latent response identification problem, fix $\mu_h = 0$ and $\psi_h^{(1)} = s$ for each item h, where the constant value s depends on the link; for example, $s = 1$ for the probit and $s = \pi/\sqrt{3}$ for the logit. The estimable parameters are then the thresholds and $\left(\psi_h^{(2)}\right)^2$, with the related ICC$_h$ = $\left(\psi_h^{(2)}\right)^2 / \left[\left(\psi_h^{(2)}\right)^2 + s^2\right]$. The point estimates and significance of the ICC$_h$ allow one to evaluate if a two-level analysis is worthwhile, and a comparison of the thresholds among the items should give some hints about possible restrictions to be imposed in the multivariate model.

2. Exploratory nonhierarchical factor analysis. To shed some light on the covariance structure of the data, it is useful to estimate the matrix of product–moment correlations among the latent responses (i.e., the polychoric correlation matrix of the items) and to use this matrix to perform an exploratory nonhierarchical (i.e., single-level) factor analysis by means of standard software.

3. Exploratory between and within factor analyses. More specific suggestions for the two-level model specification can be obtained from separate exploratory factor analyses on the estimated between and within correlation matrices of the latent responses. The results of this two-stage procedure are expected to be similar to those obtained from the full two-level analysis, as in the continuous case (Longford & Muthén, 1992). The decomposition of the latent response correlation matrix into the between and within components can be obtained by means of a multivariate two-level ordinal model with unconstrained covariance structure. For each item, the equation for the latent response is just Equation 15, but now the items are jointly modeled with an unconstrained between covariance matrix $Var\{(e_{1j}^{(2)}, \cdots, e_{Hj}^{(2)})'\}$ and an unconstrained within covariance matrix $Var\{(e_{1ij}^{(1)}, \cdots, e_{Hij}^{(1)})'\}$. Despite the latent nature of the involved variables, the correlation matrices are identified. Note that the number of random effects in this multivariate model is equal to twice the number of considered items, so the estimation process is computationally demanding. If the computation takes too long, it may be advisable to consider an approximate solution, assigning a score to the item categories and fitting a multivariate two-level model for continuous responses: The resulting correlation matrices will have slightly attenuated values, unless the distributions of the transformed variables are very far from the Gaussian distribution (Muthén & Kaplan, 1985).

4. Confirmatory two-level factor analysis. The results of the exploratory two-stage factor analysis outlined in Step 3 are used to specify one or more confirmatory two-level ordinal factor models as defined by Equation 4. These models can be fitted by means of likelihood or Bayesian methods and compared on the basis of appropriate indicators. The exploratory two-stage factor analysis of Step 3 provides fine initial values for the chosen estimation procedure, which may allow a substantial gain in computational time. Note that a large amount of computational

time can be saved by omitting the cluster-level item-specific errors $e_{hj}^{(2)}$, so that the variances of the subject errors $e_{hij}^{(1)}$ are in fact the total specificities. As illustrated earlier, this simplification prevents a full variance decomposition and the computation of the related quantities, but it is expected to be of minor importance because the interest of the researcher centers on the factor structure.

APPLICATION

The ordinal multilevel factor model was used to analyze five items on job satisfaction taken from a telephone survey conducted on 1998 graduates of the University of Florence, from 1 to 2 years after they obtained their degree.

The question on job satisfaction was asked to the employed graduates. Altogether the considered data set includes 2,432 graduates from 36 degree programs, with a highly unbalanced structure: The minimum, median, and maximum number of employed graduates per degree program are 3, 31.5, and 495, respectively.

The question—How satisfied are you with the following aspects of your present job?—required a response on a 5-point scale ranging from 1 (*very much satisfied*) to 5 (*very unsatisfied*). The five considered items were earnings, career (career opportunities), consistency (consistency with degree program curriculum), professionalism (acquisition of professionalism), and interests (correspondence with one's own cultural interests). The univariate distributions of the items are reported in Table 1. Note that the number of responses for each item is different due to item nonresponse. The multilevel factor model adopted here allows for missing item values: ML estimates are consistent under the usual missing at random (MAR) assumption (Little & Rubin, 2002).

TABLE 1
Univariate Distributions of Job Satisfaction Items. Graduates From the University of Florence, 1998

	Level of Satisfaction					*Total*	
Item	*1*	*2*	*3*	*4*	*5*	*%*	*N*
1. Earnings	7.8	23.9	38.1	20.5	9.7	100.0	2,421
2. Career	11.0	28.2	32.6	18.0	10.2	100.0	2,393
3. Consistency	24.5	27.5	24.2	12.5	11.3	100.0	2,427
4. Professionalism	26.0	40.3	22.8	7.7	3.2	100.0	2,420
5. Interests	21.5	32.7	28.2	10.8	6.8	100.0	2,419
Total	18.2	30.5	29.2	13.9	8.2	100.0	12,080

The main aim of the analysis is to describe and summarize the aspects of satisfaction measured by the five considered items, separately for the graduate and degree program levels. The two-level factor model for ordinal variables, defined by Equations 1 and 4, is a useful tool to achieve this goal. The model is quite complex and, whichever algorithm is used, the fitting process is very time consuming, so it is advisable to follow the exploratory steps outlined earlier.

Univariate Two-Level Models

The analysis begins by fitting the univariate ordinal random intercept models (Equation 15), using the logit link for consistency with the confirmatory factor model. The ML estimates, via adaptive quadrature, are reported in Table 2.

The between proportion of variance of the latent responses, ICC_h, is significantly different from zero for all items, as shown by the Likelihood Ratio Test (LRT) comparing the models with and without random intercept. Note that when the LRT is testing on the boundary of the parameter space, as in this case, the limiting distribution of the LRT statistic is not the usual χ_1^2, but instead a 50–50 mixture of a χ_0^2 (i.e., a point mass at zero) and a χ_1^2. Therefore the p values reported in Table 2 are halved (Snijders & Bosker, 1999).

The estimated ICC is low for the last two items (1.8% and 2.1%) and it is mild for the first three items, ranging from 5.5% to 8.7%. Such values of the ICC, although mild in terms of the latent response, imply relevant variations in the probabilities of the observed responses for different clusters.

A comparison among the thresholds gives some ideas on possible constraints on the thresholds in the multivariate model. In particular, the differences between adjacent thresholds among the items should be compared to informally evaluate the plausibility of the equal latent thresholds structure outlined earlier. In this case,

TABLE 2
Univariate Ordinal Logit Random Intercept Models: ICC, Thresholds, and LRT Comparing Models With and Without Random Intercept. Graduates From the University of Florence, 1998

		Thresholds				*LRT*	
Item	*ICC (%)*	γ_1	γ_2	γ_3	γ_4	*Statistic*	*p*
1. Earnings	5.5	–2.56	–0.81	0.86	2.29	88.40	.000
2. Career	8.7	–2.35	–0.63	0.84	2.14	145.68	.000
3. Consistency	6.6	–1.14	0.11	1.25	2.19	98.80	.000
4. Professionalism	1.8	–1.03	0.71	2.15	3.46	6.18	.006
5. Interests	2.1	–1.24	0.24	1.63	2.71	17.79	.000

Note. ICC = intraclass correlation coefficient; LRT = Likelihood Ratio Test.

these differences are similar for all the items, except for the third one, which shows smaller differences. This suggests that the third item has a higher variability, as also confirmed by the variances calculated after item scoring (seen later in Table 5). The confirmatory factor analysis includes a test comparing the equal latent thresholds structure with the unconstrained one.

Exploratory Nonhierarchical Factor Analysis

The second step requires the estimation of the matrix of product–moment correlations among the latent responses; that is, the polychoric correlation matrix (see Table 3) with entries that are all significantly different from zero. This matrix is used to perform an exploratory ML factor analysis via standard software. The results of this analysis (Table 4) suggest the presence of two factors: a cultural factor (labeled Factor 1) that explains primarily the consistency–professionalism–interests correlations, and a status factor (labeled Factor 2), explaining mainly the earnings–career correlation. Given the low proportions of between variance (ICCs of Table 2),

TABLE 3
Polychoric Correlation Matrix of the Items,
Graduates From the University of Florence, 1998

Item	*1*	*2*	*3*	*4*	*5*
1. Earnings	1.00				
2. Career	0.54	1.00			
3. Consistency	0.11	0.25	1.00		
4. Professionalism	0.28	0.45	0.54	1.00	
5. Interests	0.16	0.33	0.61	0.58	1.00

TABLE 4
Exploratory Factor Analysis on the Polychoric Correlation Matrix
of the Items: Varimax Rotated Factors and Communalities. Graduates
From the University of Florence, 1998

	Factor Loadings		
	Factor 1	*Factor 2*	*Communality*
1. Earnings	.08	.65	.43
2. Career	.26	.80	.70
3. Consistency	.77	.07	.60
4. Professionalism	.68	.34	.58
5. Interests	.78	.16	.63

this structure is expected to be quite similar to the within structure, although it may be very different from the between structure.

Exploratory Between and Within Factor Analyses

The third step of the analysis calls for the decomposition of the overall correlation matrix of the latent responses into the between and within components. This task would require fitting a two-level multivariate ordinal model with five random effects for each level, which takes too long to be fitted with numerical integration. Therefore an approximate procedure is adopted, assigning a score to each item category. Various sophisticated scoring systems could be applied (Fielding, 1999), but given the preliminary nature of this step, the simplest scoring system is applied, assigning integer values 1 to C_h to the categories. After scoring, the within and between covariance matrices can be estimated by fitting a multivariate two-level model for continuous responses. To this end, the MLwiN software with RIGLS algorithm (Rasbash, Steele, Browne, & Prosser, 2004) is used. RIGLS yields restricted ML estimates, which are less biased for variance–covariance parameters than full ML (Goldstein, 2003).

The results are shown in Tables 5 and 6. For Table 5, note the following points: (a) It is clear from the last row that the third item (consistency) has the higher variability, as already noted in the univariate analysis (Table 2); (b) the between percentages of variance (i.e., the values on the diagonal) are in line with the ICCs of Table 2; and (c) the between percentages tend to be higher for covariances than for variances.

For Table 6, note the following points: (a) The total correlation matrix, which is obtained from the between and within components, is similar to the polychoric correlation matrix (Table 3), with a moderate attenuation; (b) the between and within correlation matrices have quite different structures (e.g., the between correlations are always higher than the within correlations); and (c) the within correla-

TABLE 5
Two-Level Multivariate Model on Item Scores: Between Percentage of Variance (in Italics) and covariance, and Total Variance of Items; Graduates From the University of Florence, 1998

Item	*1*	*2*	*3*	*4*	*5*
1. Earnings	*5.90*				
2. Career	12.96	*8.63*			
3. Consistency	21.02	13.09	*7.37*		
4. Professionalism	10.55	7.63	6.62	*2.30*	
5. Interests	9.57	4.75	6.07	2.76	*2.36*
Total variance	1.15	1.31	1.68	1.04	1.31

TABLE 6
Two-Level Multivariate Model on Item Scores: Correlation Matrix Decomposition. Graduates From the University of Florence, 1998

Item	*1*	*2*	*3*	*4*	*5*
Between					
1. Earnings	1.00				
2. Career	.89	1.00			
3. Consistency	.36	.40	1.00		
4. Professionalism	.72	.69	.79	1.00	
5. Interests	.39	.32	.81	.62	1.00
Within					
1. Earnings	1.00				
2. Career	.46	1.00			
3. Consistency	.10	.23	1.00		
4. Professionalism	.23	.39	.48	1.00	
5. Interests	.14	.31	.55	.52	1.00
Total					
1. Earnings	1.00				
2. Career	.49	1.00			
3. Consistency	.11	.24	1.00		
4. Professionalism	.25	.40	.49	1.00	
5. Interests	.15	.30	.55	.53	1.00

tion matrix is similar to the total correlation matrix, due to the low proportion of between variances and covariances.

The results of the exploratory ML factor analyses performed on the within and between correlation matrices of Table 6 are reported in Table 7.

As for the within structure, Bartlett's test indicates that two factors are sufficient (p value = .5082). The factor loadings are similar to those found in the non-hierarchical analysis (Table 4).

As for the between structure, although one factor is not enough, the estimation with two or more factors encounters a Heywood case. We decided to retain two factors, forcing the specificities to be nonnegative. The second factor is measured by all items, whereas the first factor has relevant loadings only for the last three items.

Confirmatory Two-Level Factor Analysis With Unconstrained Thresholds

Finally, in light of the results of the exploratory analysis, a two-level confirmatory factor analysis is performed using the model defined by Equations 1 and 4. ML estimates are obtained with M*plus* (Muthén & Muthén, 2006). M*plus* performs ML

TABLE 7
Exploratory Maximum Likelihood Factor Analysis on the Within and Between Correlation Matrices Estimated From Item Scores: Factor Loadings and Communalities (Varimax Rotation Only for the Within Part), Graduates From the University of Florence, 1998

	Within			*Between*		
	Loadings			*Loadings*		
Item	*F1*	*F2*	*Communality*	*F1*	*F2*	*Communality*
1. Earnings	.07	.59	.35	.00	1.00	1.00
2. Career	.25	.75	.63	.08	.89	.80
3. Consistency	.72	.07	.53	.93	.36	1.00
4. Professionalism	.64	.32	.50	.57	.72	.84
5. Interests	.74	.16	.58	.71	.39	.66

estimation via an EM algorithm, solving the integrals with adaptive Gaussian quadrature. M*plus* version 4 allows modeling of ordinal responses using the probit or the logit link. However, M*plus* version 3 was used here, which employs only the logit link. This implies a little change with respect to the model discussed earlier, namely the distribution of $e_{hij}^{(1)}$ is logistic instead of Gaussian. Note that the logit implies that $\psi_h^{(1)}$ is equal to $\pi/\sqrt{3}$, instead of 1 for the probit, whereas $\mu_h = 0$ as for the probit.

The within and between structures emerging from the exploratory analyses are not equally reliable: The within part is estimated on a large number of observations and Bartlett's test clearly indicates the presence of two factors, whereas the between part is estimated on only 36 degree programs and the estimation is complicated by the presence of a Heywood case.

Therefore, for the within part of the model, the two-factor structure suggested by the exploratory within factor analysis (see Table 7) is retained, constraining to zero the loadings that were close to zero, that is the loading of earning in the first factor and the loadings of consistency and interests in the second factor. As for the between structure, because the hints from the exploratory analysis are less clear, two configurations at this level have been tried: (a) a one-factor unconstrained structure (Model M1), and (b) a two-factor structure (Model M2), with unconstrained loadings in the first factor and two loadings equal to zero in the second factor (earning and career; see Table 7).

Models M1 and M2 are fitted without imposing any restriction on the thresholds, to preserve the covariance structure from possible misspecifications of the thresholds. Because we are not particularly interested in decomposing the item specificities, to reduce the computational effort the cluster-level item-specific errors $e_{hj}^{(2)}$ are omitted, so the specificities are in fact total specificities. The models are fitted using five quadrature points for each factor, with a total of 125 and 625 quadrature points for M1 and M2, respectively. Some limited trials suggest that larger numbers of quadrature points do not improve the estimates in a significant manner.

The LRT comparing the models M1 and M2 clearly indicates that the second model is better. The preferred model, M2, has 35 estimable parameters: 20 thresholds $\gamma_{c,h}$, 5 factor loadings $\lambda_{mh}^{(1)}$ and 2 factor standard deviations $\sigma_m^{(1)}$ at the graduate level ($m = 1, 2$); and 6 factor loadings $\lambda_{mh}^{(2)}$ and 2 factor standard deviations $\sigma_m^{(2)}$ at the degree program level ($m = 1, 2$). The parameter estimates for Model M2 are reported in Table 8. The interesting part of the model is the covariance structure at both levels, which does not depend on the item means and thresholds and can be summarized by the communalities (see Table 9). These values are obtained as suitable transformations of model parameters: The factor-specific communalities are computed from formulas such as Equation 14; the total communality is obtained by summing the row values FW1, FW2, FB1, and FB2 (see Equation 13); and the last column of Table 8 is the percentage of total communality due to the between level. The following points should be noted: (a) For the first three items, the between component is greater for the communality (last column of Table 9) than for the total variance (ICCs of Table 2); (b) the last two items, professionalism and interests, are very poorly explained by the factors at degree program level; and (c) the first factor at the degree program level, FB1, is interpretable as a status factor, whereas the second one, FB2, is essentially related to the item consistency.

TABLE 8
Confirmatory Two-Level Factor (F) Analysis (Model *M*2): Parameter Estimates, Graduates From the University of Florence, 1998

	Loadings				*Thresholds*			
	Within (W)		*Between (B)*					
Item	$\lambda_{1h}^{(1)}$	$\lambda_{2h}^{(1)}$	$\lambda_{1h}^{(2)}$	$\lambda_{2h}^{(2)}$	$\gamma_{1,h}$	$\gamma_{2,h}$	$\gamma_{3,h}$	$\gamma_{4,h}$
1. Earnings	—	1.09	0.75	—	–3.96	–1.39	1.24	3.29
2. Career	1[a]	1[a]	1[a]	—	–3.64	–0.92	1.46	3.37
3. Consistency	2.30	—	0.32	1[a]	–1.80	0.19	1.98	3.39
4. Professionalism	2.25	0.41	0.31	0.27	–1.79	1.07	3.36	5.27
5. Interests	2.85	—	0.09	0.47	–2.34	0.34	2.83	4.62
Factor variance	0.75	3.09	0.71	0.45				

[a]Denotes a fixed value.

TABLE 9
Confirmatory Two-Level Factor (F) Analysis (Model *M2*): Communalities, Graduates From the University of Florence, 1998

	% Communality					
	Within (W)		*Between (B)*			
Item	*FW1*	*FW2*	*FB1*	*FB2*	*Total*	*% Between on Total*
1. Earnings	—	49.9	5.4	—	55.2	9.7
2. Career	9.6	39.4	9.0	—	58.0	15.6
3. Consistency	51.1	—	0.9	5.7	57.7	11.5
4. Professionalism	49.2	6.7	0.9	0.4	57.2	2.3
5. Interests	64.2	—	0.1	1.0	65.3	1.6

The factor scores at degree program level are represented in Figure 1, where the labels are attached to the degree programs with extreme scores: The points on the right side of the figure indicate a high satisfaction on earning and career, whereas the points on the top denote a high satisfaction on consistency. Note that there are two degree programs in the lower left corner (philosophy and natural sciences) with low satisfaction on both dimensions.

The analysis could be deepened by adding some covariates, but this is beyond the goal of this article. This extension is straightforward even if it might require a substantial increase in computational time.

Finally, a quick test to evaluate if the model selection is influenced by the omission of the cluster-level specific errors $e_{hj}^{(2)}$ consists in fitting the same models as before except for treating the responses as continuous (i.e., using the item scores). In such a case, M*plus* avoids numerical integration, so estimation takes only a few seconds. Two models on item scores are fitted: M1* and M2*, with the same factor structure as M1 and M2, respectively. In both M1* and M2* the cluster-level item-specific errors $e_{hj}^{(2)}$ are omitted. The LRT comparing these two models confirms that M2* is better than M1* (LR statistic = 51.7, *df* = 3). Subsequently, the cluster-level item-specific errors $e_{hj}^{(2)}$ are added to M1* and M2*. Denoting with M1+ and M2+ the resulting models, the LRT comparing M1+ with M2+ leads to a less clear result (LR statistic = 7.4, *df* = 3, *p* = .06). Therefore the choice of not decomposing the specificities may have unexpected consequences on the selection of the factor structure.

Confirmatory Two-Level Factor Analysis With Constrained Thresholds

In the search for a more parsimonious specification of the model, it is interesting to consider the equal latent thresholds structure described earlier, comparing Model

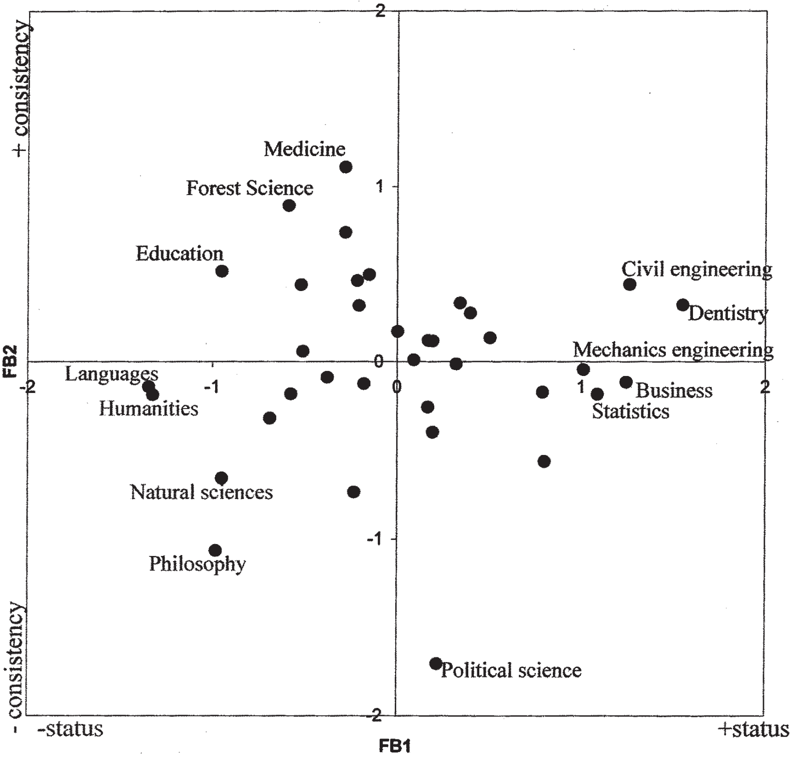

FIGURE 1 Estimated factor scores for the degree programs (model $M2$). Graduates from the University of Florence, 1998.

M2 with a model having the same factor structure, but constrained thresholds. This structure cannot be easily imposed in M*plus* because the subject-level item-specific standard deviations $\psi_h^{(1)}$ are assumed to be equal across items and cannot be defined as model parameters. This point can be overcome in two ways.

First, it is possible to add a set of fictitious factors each pointing to one item, except for the reference item. This allows definition of the item-specific means μ_h and specificities $\psi_h^{(1)}$. However, this solution is computationally inefficient, as it increases by $(H-1)$ the number of latent variables and thus the dimension of the integration in the marginal likelihood. Second, it is possible to impose nonlinear constraints on the thresholds to obtain the desired threshold structure. As previously noted, under the assumption that the thresholds $\gamma_{c,h}$ are constant across items and hence written as γ_c, the actual thresholds are $\tau_{c,h} = (\gamma_c - \mu_h)/\psi_h^{(1)}$, for a total of $(C-1) + 2 \cdot (H-1)$ free parameters. In the M*plus* parametrization μ_h and $\psi_h^{(1)}$ are assumed to be constant across items, so the threshold model for the ordinal variables is characterized by $(C-1)\cdot H$ estimable thresholds: therefore $(C-3)\cdot(H-1)$ constraints must be imposed to get the correct number of free parameters. To this end, note that the relation $\tau_{c,h}=(\gamma_c-\mu_h)/\psi_h^{(1)}$ implies the following equalities for any $c = 2, \ldots, C-1$, and any pairs of items h and h^*:

$$\frac{\tau_{c,h}-\tau_{1,h}}{\tau_{c,h^*}-\tau_{1,h^*}}=\frac{\psi_{h^*}^{(1)}}{\psi_h^{(1)}}.$$

Hence, in the present case ($C = 5$ and $H = 5$) the required $2\cdot 4 = 8$ constraints on the actual thresholds could be:

$$\frac{\tau_{3,h}-\tau_{1,h}}{\tau_{3,h^*}-\tau_{1,h^*}}=\frac{\tau_{2,h}-\tau_{1,h}}{\tau_{2,h^*}-\tau_{1,h^*}}$$

$$\frac{\tau_{4,h}-\tau_{1,h}}{\tau_{4,h^*}-\tau_{1,h^*}}=\frac{\tau_{2,h}-\tau_{1,h}}{\tau_{2,h^*}-\tau_{1,h^*}}$$

for $h^* = 2$ (the reference item) and $h = 1, 3, 4, 5$.

After estimation, the item-specific standard deviations and means (with respect to the reference item) can be recovered using appropriate formulas; for example,

$$\frac{\psi_h^{(1)}}{\psi_{h^*}^{(1)}}=\frac{\tau_{2,h^*}-\tau_{1,h^*}}{\tau_{2,h}-\tau_{1,h}}$$

and

$$\begin{aligned}\mu_h-\mu_{h^*} &= \psi_{h^*}^{(1)}\cdot\tau_{1,h^*}-\psi_h^{(1)}\cdot\tau_{1,h}\\ &= \psi_{h^*}^{(1)}\cdot\left[\tau_{1,h^*}-\frac{\psi_h^{(1)}}{\psi_{h^*}^{(1)}}\cdot\tau_{1,h}\right].\end{aligned}$$

M*plus* is used to fit a model with the same factor structure of Model M2, but with equal latent thresholds. This leads to a model with 35 parameters and 8 nonlinear constraints on the thresholds. The LRT statistic for the equal latent thresholds assumption is 17.19 ($df = 8$, $p = .028$). Therefore, with the data at hand the use of such structure is questionable, although the consequences on the communalities and factor scores are found to be modest.

An alternative software for fitting the models described in this article is the `gllamm` command of Stata (Rabe-Hesketh et al., 2004b), a highly flexible procedure that allows fitting of the two-level factor model for ordinal variables both with unconstrained thresholds and with equal latent thresholds. In `gllamm` the equal latent thresholds structure is appealing, as it is implemented through a special link function, the scaled ordered probit (option `link(soprobit)`), whose scale parameters are the standard deviations of the specificities, $\psi_h^{(1)}$.

The `gllamm` command performs ML, using a Newton–Raphson algorithm with adaptive Gaussian quadrature. In our application, the results obtained with `gllamm` are similar to those yielded by M*plus*, although the computational times are substantially longer.

CONCLUDING REMARKS

Multilevel factor models for ordinal variables are useful but complex tools, giving rise to problems of specification, identification, estimation, and interpretation. At present the major obstacle to a wide use of such models is software limitations. To our knowledge the only widespread packages able to yield full information ML estimates for the models discussed here are M*plus* and the `gllamm` command of Stata.

Full information ML has several advantages over limited information methods, namely (a) efficiency; (b) availability of a likelihood, allowing likelihood-based inference; and (c) ability to deal with many patterns of missingness, yielding consistent estimates under the usual MAR assumption. The techniques to obtain full information ML can be classified along two dimensions (Rijmen et al., 2003): the method of numerical integration of the intractable integrals used to approximate the marginal likelihood and the type of algorithm used to maximize the approximate marginal likelihood.

Numerical integration can be deterministic, such as Gaussian quadrature (adaptive or not), or stochastic, such as Monte Carlo integration. In general the computational time of Gaussian quadrature is roughly proportional to the product of the number of quadrature points for all the latent variables used, so models with three or more factors per level may take too much time to be of practical use. In such cases, Monte Carlo integration may be more convenient. There are promising attempts to improve the efficiency of numerical integration techniques (e.g., spherical quadrature: Rabe-Hesketh et al., 2005; quasi-Monte Carlo: Pan & Thompson, 2004).

The maximizing algorithm can perform the maximization directly on the marginal likelihood, such as the Newton–Raphson, or indirectly on some variant of the likelihood, such as the EM. Further research is needed to assess the relative merits of Newton–Raphson, EM, and their numerous variants, and to assess the interactions with the numerical integration techniques.

A promising route is the adaptation of simulation-based methods (Gouriéroux & Monfort, 1996) to the class of multilevel factor models: An interesting example in this respect is the application of Mazzolli (2001) concerning a multilevel structural equation model with ordinal variables. Moreover, in the search for approximate but computationally efficient methods, the development of limited information methods (Muthén & Satorra, 1995) may be worthwhile.

In the Bayesian paradigm there is a growing research activity aimed at developing efficient MCMC algorithms for models with latent variables: In particular, Ansari and Jedidi (2000) and Goldstein and Browne (2005) treated multilevel factor models with binary responses, and Fox and Glas (2002) considered more general multilevel structural models.

Although faster estimation algorithms can be developed, the supplementary computational effort needed to treat the response variables as ordinal, instead of continuous, is inevitably not negligible, so one can legitimately wonder whether the effort is adequately repaid in terms of the quality of statistical inference. A general answer is obviously not possible. The results of Muthén and Kaplan (1985) suggest that, in standard factor models, treating the ordinal variables as continuous is not severely harmful when the frequency distributions are unimodal with an internal mode. However, the use of a proper model is always a desirable feature of the analysis and the resulting inferences are generally more reliable.

As a final note, multilevel factor models should be handled with care. Hence, even if very efficient estimation algorithms were available, it is a good practice, especially in the case of categorical response variables, to fit a multilevel factor model as the final step of the analysis, after having explored the data with simpler techniques.

REFERENCES

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis* (3rd ed.). New York: Wiley.

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

Fielding, A. (1999). Why use arbitrary point scores?: Ordered categories in models of educational progress. *Journal of the Royal Statistical Society, A162,* 303–328.

Fox, J. P., & Glas, C. A. W. (2002). Modeling measurement error in structural multilevel models. In G. A. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 243–267). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Arnold.

Goldstein, H., & Browne, W. J. (2002). Multilevel factor analysis modelling using Markov Chain Monte Carlo (MCMC) estimation. In G. A. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 225–242). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Goldstein, H., & Browne, W. J. (2005). Multilevel factor analysis models for continuous and discrete data. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 453–475). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Gouriéroux, C., & Monfort, A. (1996). *Simulation based econometric methods.* New York: Oxford University Press.

Grilli, L., & Rampichini, C. (2003). Alternative specifications of multivariate multilevel probit ordinal response models. *Journal of Educational and Behavioral Statistics, 28,* 31–44.

Jöreskog, K. G. (2005). *Structural equation modeling with ordinal variables using LISREL.* Retrieved February10, 2005, from http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf

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

Jöreskog, K. G., & Sörbom, D. (2005). *LISREL for Windows.* Lincolnwood, IL: Scientific Software International.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: Wiley.

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

Mazzolli, B. (2001, April). *A multilevel structural equation model with polytomous and dichotomous data.* Paper presented at the Third International Conference on Multilevel Analysis, Amsterdam.

Moustaki, I., Jöreskog, K. G., & Mavridis, D. (2004). Factor models for ordinal variables with covariate effects on the manifest and latent variables: A comparison of LISREL and IRT approaches. *Structural Equation Modeling, 11,* 487–513.

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

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

Muthén, B. O., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38,* 171–189.

Muthén, B. O., & Satorra, A. (1995). Technical aspects of Muthén's LISCOMP approach to estimation of latent variables relations with a comprehensive measurement model. *Psychometrika, 60,* 489–503.

Muthén, L. K., & Muthén, B. O. (2006). *M*plus *user's guide* (4th ed.). Los Angeles: Muthén & Muthén.

Pan, J., & Thompson, R. (2004). Quasi-Monte Carlo estimation in generalized linear mixed models. In A. Biggeri, E. Dreassi, C. Lagazio, & M. Marchi (Eds.), *Proceedings of the 19th International Workshop on Statistical Modelling* (pp. 239–243). Firenze, Italy: Firenze University Press.

Rabash, J., Steele, F., Browne, W. J., & Prossner, B. (2004). *A user's guide to MLwiN version 2.0.* London: Institute of Education.

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

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004b). *GLLAMM manual* (U.C. Berkeley Division of Biostatistics Working Paper Series No. 160). Berkeley: University of California.

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.

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

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

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modelling.* London: Sage.