MULTIVARIATE BEHAVIORAL RESEARCH, *40*(3), 281–301

Mixed-Effects Logistic Regression Models for Indirectly Observed Discrete Outcome Variables

Jeroen K. Vermunt
Department of Methodology and Statistics
Tilburg University

A well-established approach to modeling clustered data introduces random effects in the model of interest. Mixed-effects logistic regression models can be used to predict discrete outcome variables when observations are correlated. An extension of the mixed-effects logistic regression model is presented in which the dependent variable is a latent class variable. This method makes it possible to deal simultaneously with the problems of correlated observations and measurement error in the dependent variable. As is shown, maximum likelihood estimation is feasible by means of an EM algorithm with an E step that makes use of the special structure of the likelihood function. The new model is illustrated with an example from organizational psychology.

In many regression applications, observations have some kind of clustering, with observation within clusters tending to be correlated. Examples include observations obtained via repeated measurements, data collected by two-stage cluster sampling designs, and hierarchical or multilevel data in which units are grouped at different levels. A well established approach to modeling clustered data introduces cluster-level random effects in the model of interest (Bryk & Raudenbush, 1992; Goldstein, 1995; Snijders & Bosker, 1999). Such models are called mixed-effects, random-effects, hierarchical, or multilevel models. Whereas most of the work on mixed-effects models is for continuous outcome variables, recently models for categorical outcome variables have received more attention. This article deals with mixed-effects models for dichotomous, ordinal, and nominal response variables or, more precisely, with mixed-effects logistic regression (MELR) models

Correspondence concerning this article should be addressed to Jeroen K. Vermunt, Department of Methodology and Statistics, Tilburg University, P. O. Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: J.K.Vermunt@kub.nl

(Agresti, Booth, Hobert, & Caffo, 2000; Hartzel, Agresti, & Caffo, 2001; Hedeker, 1999, 2003; Hedeker & Gibbons, 1996; Skrondal & Rabe-Hesketh, 2004; Wong & Mason, 1985).

Existing MELR models assume that the discrete dependent variable is an observable variable, but in many social and behavioral sciences applications it is impossible to measure a variable directly. Examples of discrete outcome variables that cannot be observed directly are the presence/absence of a mental or behavioral disorder (depressed/not depressed), personality types, developmental stages, consumer segments, etcetera. In such situations, one will typically have several imperfect indicators (items) that should be combined into a single scale or typology. An elegant method to construct such discrete latent classifications from multiple indicators is latent class (LC) analysis (Goodman, 1974; Hagenaars, 1990; McCutcheon, 1987).

In this article, I propose a model that combines a MELR model with a LC type structure for the (latent) dependent variable. This LC-MELR model makes it possible to simultaneously address the problems of dependent observations and measurement error in the outcome variable. The presented approach is especially useful in research settings in which it makes more sense to treat the latent variable of interest as discrete rather than as continuous, a situation that often occurs in social and behavioral research applications.

The proposed approach is related to the work of Raudenbush and Sampson (1999), Fox and Glas (2001), and Goldstein and Browne (2002). A difference is that these authors assume that the latent dependent variable is a continuous instead of a discrete variable. The regression part of their models has, therefore, the form of a standard mixed-effects linear model, and the measurement part has the form of an item response theory (IRT) model. Another difference is that here we work within a maximum likelihood (ML) framework, whereas these authors use either approximate ML or Bayesian estimation procedures. In order to make ML estimation of the LC-MELR model feasible, I propose an adapted version of the EM algorithm that makes use of the conditional independence assumptions implied by the LC-MELR model.

In the next section, I motivate the LC-MELR using an empirical example from organization psychology. Then, I describe the various components of this model and discuss estimation issues that are specific for this new model. Subsequently, the results obtained when applying the model to the empirical example are presented. The article ends with a short discussion.

INTRODUCTION OF THE TASK VARIETY APPLICATION

The empirical application I use to illustrate the proposed method comes from a Dutch study on the effect of team characteristics on individual work conditions

(Van Mierlo, 2003). A questionnaire was filled in by 886 employees from 88 teams of two organizations, a nursing home and a domiciliary care organization. Various aspects of the work condition were assessed, one of which was the perceived task variety, which was measured by five dichotomous items. Besides these five questionnaire items, there is information on various individual-level covariates.

The model proposed in this article is useful in situations in which the researcher desires to build a latent typology or taxonomy from a set of observed item responses. Thus, rather than working with a continuous underlying latent variable as in factor analysis and latent trait models, the purpose should be to construct a discrete typology. For this purpose, we can use latent class analysis, a method that is becoming more popular in social and behavioral research. Often, it is more natural to summarize the information on multiple items, indicators or symptoms into a discrete typology instead of a continuous scale, especially in a diagnostic context. Examples are instruments to determine presence of a mental or behavioral disorder, personality type, developmental stage of a child, mastering of certain school tasks, and presence of certain unfavorable work conditions.

The second element of the proposed method is that it is suited for the analysis of nested data structures: for example, children belonging to the same family or the same school, patients treated at the same clinic, consumers living in the same neighborhood, or employees working in the same team. More specifically, a multilevel regression model with random effects is used to take the dependencies between observations from the same group into account. This technique can also be used to build multilevel explanations for individual-level outcome variables and to determine the relative importance of group-level and individual-level factors in the prediction of these outcome variables.

The purpose of the analysis of the task variety data is twofold. On one hand, we desire to construct a diagnostic instrument for task variety yielding a discrete classification into two—or possibly three—categories. On the other hand, we are interested in detecting whether there are team differences in the perceived individual task variety, that is, whether the lack of variation in the work is systematic within certain teams or whether it mainly depends on individual characteristics. The individual-level covariates are used to describe the within-group differences, as well as to correct for possible composition effects.

Using certain simplifying assumptions, the analysis could be performed with standard multilevel logistic regression methods. A simple sum score with a certain cutoff point could be use to obtain a two-category classification of task variety. Disadvantages of such a procedure are among others that the cutoff point is always somewhat arbitrary in the sense that it can be anything, that measurement error is not taken into account, and that it cannot easily be used with missing values on the items. Another option might be to perform a standard latent class analysis without taking the nesting into account, and use the latent class assignments as an observed outcome variable in a subsequent (multilevel) regression analysis. Such a two-step procedure has two important disadvantages: it introduces measurement errors leading to attenuated effects and it does not take the nested data structure into account when building the latent typology, which may result in biased parameter estimates and test results.

LATENT CLASS MIXED-EFFECTS LOGISTIC REGRESSION MODEL

Model for an Observed Nominal Outcome Variable

Let x_{ij} denote the response on a dichotomous or nominal dependent variable of individual or level-1 unit i within cluster or level-2 unit j. A particular response is denoted by t, and the number of possible responses by T. The total number of level-2 units is denoted by J and the number of level-1 units within level-2 unit j by n_j. Let $\mathbf{w}_{ij}$ and $\mathbf{z}_{ij}$ denote the design vectors associated with S fixed and R random effects, respectively.

The MELR model proposed by Hedeker (1999, 2003) is defined as

$$P\left(x_{ij}=t \mid \mathbf{w}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\beta}_j\right)=\frac{\exp\left(\eta_{ijt}\right)}{\sum_{t=1}^{T}\exp\left(\eta_{ijt}\right)}, \tag{1}$$

with

$$\eta_{ijt}=\mathbf{w}_{ij}'\boldsymbol{\alpha}_t+\mathbf{z}_{ij}'\boldsymbol{\beta}_{jt}.$$

Here, $\boldsymbol{\alpha}_t$ is the vector of unknown fixed effects corresponding to response category t, and $\boldsymbol{\beta}_{jt}$ is the vector of unknown random effects for level-2 unit j and response category t. It should be noted that, as in standard multinomial logistic regression analysis, with T response categories, only $T - 1$ sets of effects can be identified. For identification purposes, we may, for example, set the fixed and random effects corresponding to the first category equal to zero—$\boldsymbol{\alpha}_1 = \boldsymbol{\beta}_{j1} = 0$—which amounts to using dummy coding with the first category as reference category. An alternative to dummy coding is effect coding, implying that parameters sum to zero across the categories of the response variable; that is,

$$\sum_{t=1}^{T}\boldsymbol{\alpha}_t=\sum_{t=1}^{T}\boldsymbol{\beta}_{jt}=0.$$

As is most common, we assume the distribution of the random effects $\boldsymbol{\beta}_{jt}$ to be multivariate normal with zero mean vector and covariance matrix $\boldsymbol{\Sigma}_t$. For parameter estimation, it is convenient to standardize and orthogonalize the random effects. For this, let $\boldsymbol{\beta}_{jt} = \mathbf{C}_t\boldsymbol{\theta}_j$, where $\mathbf{C}_t\mathbf{C}_t' = \boldsymbol{\Sigma}_t$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_t$. The reparameterized model is then

$$\eta_{ijt} = \mathbf{w}_{ij}'\boldsymbol{\alpha}_t + \mathbf{z}_{ij}'\mathbf{C}_t\boldsymbol{\theta}_j. \quad (2)$$

Hedeker's formulation of the MELR model assumes that the random effects belonging to the different categories of the dependent variable are perfectly correlated. More specifically, the same random effects $\boldsymbol{\theta}_j$ are scaled in a different manner for each t by the unknown $\mathbf{C}_t$. This formulation, which is equivalent to Bock's nominal response model (Bock, 1972), is based on the assumption that each nominal category is related to an underlying latent response tendency.

Another formulation of the MELR is

$$\eta_{ijt} = \mathbf{w}_{ijt}'\boldsymbol{\alpha} + \mathbf{z}_{ijt}'\mathbf{C}\boldsymbol{\theta}.$$

In this specification, which is more common in the econometric literature, the design vectors may be category specific whereas the parameters vector need not be category specific (Hartzel et al., 2001).

Indirectly Observed Nominal Dependent Variable

Suppose that the outcome variable x_{ij} cannot be observed directly, but only indirectly by means of a set of K categorical items. Let y_{ijk} denote the response of individual i within cluster j on item k. A particular level of item k is denoted by d_k and its number of categories by D_k. For the response variable of interest, we keep the notation of the previous subsection. A difference is, of course, that it is now a latent instead of an observed variable.

The indirectly observed response variable is related to the item responses by means of a LC model (Goodman, 1974; Hagenaars, 1990; McCutcheon, 1987). Let $P(y_{ijk} = d_k | x_{ij} = t)$ denote the probability that individual i of cluster j gives response d_k on item k given that (s)he belongs to latent class t. The basic assumption of the LC model is that the observed item responses are mutually independent given class membership. If we combine the MELR model with a LC model, we obtain the following probability structure for the joint conditional distribution of x_{ij} and $\mathbf{y}_{ij}$:

$$P\left(x_{ij} = t, \mathbf{y}_{ij} = \mathbf{d} \mid \mathbf{w}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\theta}_j\right) = P\left(x_{ij} = t \mid \mathbf{w}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\theta}_j\right)\prod_{k=1}^{K} P\left(y_{ijk} = d_k \mid x_{ij} = t\right). \quad (3)$$

Here, the product

$$\prod_{k=1}^{K} P\left(y_{ijk} = d_k \mid x_{ij} = t\right),$$

defines the local independence structure of the LC model. The term $P(x_{ij} = t \mid \mathbf{w}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\theta}_j)$ is restricted by the MELR model defined in Equation 1. The response probabilities $P(y_{ijk} = d_k \mid x_{ij} = t)$ can be parametrized by logistic functions of the form

$$P\left(y_{ijk} = d_k \mid x_{ij} = t\right) = \frac{\exp\left(\delta_{d_k} + \gamma_{d_k t}\right)}{\sum_{d_k=1}^{D_k} \exp\left(\delta_{d_k} + \gamma_{d_k t}\right)}.$$

As is shown below, such a linear logistic parametrization will prove useful in more advanced models.

Collapsing Equation 3 over the T categories of the discrete latent variable yields the marginal conditional distribution of the observed variables $\mathbf{y}_{ij}$,

$$P\left(\mathbf{y}_{ij} = \mathbf{d} \mid \mathbf{w}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\theta}_j\right) = \sum_{t=1}^{T} P\left(x_{ij} = t, \mathbf{y}_{ij} = \mathbf{d} \mid \mathbf{w}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\theta}_j\right). \tag{4}$$

The model described in Equation 3 is not only an extension of the MELR model, it is also an extension of the concomitant variable LC model (Bandeen-Roche, Miglioretti, Zeger, & Rathouz, 1997; Dayton & Macready, 1988). This is a LC model in which class membership is regressed on covariates using a logistic regression model. The new element of our approach is that this logistic regression model for the latent classes can now contain random effects and can thus also be used with clustered observations.

MAXIMUM LIKELIHOOD ESTIMATION

Log-likelihood Function

The parameters of the LC-MELR model described in the previous section can be estimated by maximum likelihood (ML). The likelihood function is based on the probability densities of the level-2 observations, denoted by $P(\mathbf{y}_j|\mathbf{w}_j,\mathbf{z}_j)$. Note that these are independent observations whereas observations within level-2 units cannot be assumed to be independent. The log-likelihood to be maximized equals

$$\log L = \sum_{j=1}^{J} \log P\left(\mathbf{y}_j \mid \mathbf{w}_j, \mathbf{z}_j\right),$$

where

$$\begin{aligned} P\left(\mathbf{y}_j \mid \mathbf{w}_j, \mathbf{z}_j\right) &= \int_{\boldsymbol{\theta}} P\left(\mathbf{y}_j \mid \mathbf{w}_j, \mathbf{z}_j, \boldsymbol{\theta}\right) f(\boldsymbol{\theta})\, d\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\theta}} \left[\prod_{i=1}^{n_j} P\left(\mathbf{y}_{ij} \mid \mathbf{w}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\theta}\right)\right] f(\boldsymbol{\theta})\, d\boldsymbol{\theta}, \end{aligned} \tag{5}$$

Notice that $\boldsymbol{\theta}$ denotes the vector of orthonormalized random effects. As can be seen, the responses of the n_j level-1 units within level-2 unit j are assumed to be independent of one another given the random effects $\boldsymbol{\theta}$. Of course, the contributions of the level-1 units have the form of the LC-MELR model described in Equations 1–4.

The integral on the right-hand side of Equation 5 can be evaluated by the Gauss-Hermite quadrature numerical integration method (Bock & Aikin, 1981; Hedeker, 1999, 2003; Stroud & Secrest, 1966), in which the multivariate normal distribution of the random effects is approximated by a limited number of M discrete points. More precisely, the integral is replaced by a summation over M quadrature points,

$$\begin{aligned} P\left(\mathbf{y}_j \mid \mathbf{w}_j, \mathbf{z}_j\right) &\approx \sum_{m=1}^{M} P\left(\mathbf{y}_j \mid \mathbf{w}_j, \mathbf{z}_j, \boldsymbol{\theta}_m\right) \pi(\boldsymbol{\theta}_m) \\ &= \sum_{m=1}^{M} \left[\prod_{i=1}^{n_j} P\left(\mathbf{y}_{ij} \mid \mathbf{w}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\theta}_m\right)\right] \pi(\boldsymbol{\theta}_m). \end{aligned} \tag{6}$$

Here, $\boldsymbol{\theta}_m$ and $\pi(\boldsymbol{\theta}_m)$ denote the quadrature nodes and weights corresponding to the (multivariate) normal density of interest. Because the R random effect are orthogonalized, the nodes and weights of the separate dimensions equal the ones of the univariate normal density, which can be obtained from standard tables (see, e.g., Stroud & Secrest, 1966). Suppose that each of the R dimensions is approximated with Q quadrature nodes. The $M = Q^R$ R-dimensional weights are then obtained by multiplying the weights of the separate dimensions. The integral can be approximated to any practical degree of accuracy by setting Q sufficiently large.

The preferred algorithm for obtaining ML estimates in LC analysis is the well-known EM algorithm. In the Appendix, I discuss the problems associated with the implementation of a standard EM algorithm in the case of the LC-MELR model and show how these can be resolved by making use of the conditional independence assumption implied by the model. The Appendix also discusses computation of standard errors, identification problems, and software for estimating the LC-MELR model.

Intraclass Correlation

A measure that is often of interest in random-effects modeling is the intraclass correlation. It is defined as the proportion of the total variance accounted for by the level-2 units, where the total variance equals the sum of the level-1 and level-2 variances. Hedeker (2003) showed how to compute the intraclass correlation in logistic regression models containing only a random intercept; that is,

$$r_{It} = \frac{\sigma_t^2}{\sigma_t^2 + \pi^2/3}. \tag{7}$$

This formula makes use of the fact that the level-1 variance can be set equal to the variance of the logistic distribution, which equals $\pi^2/3 \approx 3.29$. The other term—σ_t^2—is the level-2 variance corresponding to the intercept of latent class t. Notice that $T - 1$ independent intraclass correlations can be computed, which means that the cluster influence is allowed to vary across contrasts between (latent) response categories.

Cluster-Specific Effects

One of the objectives of random-effects modeling may be to obtain estimates of the cluster-specific parameters. A simple estimator for $\boldsymbol{\beta}_{jt}$ is the expected "a posteriori" (EAP), posterior mean, or empirical Bayes estimator $\bar{\boldsymbol{\beta}}_{jt}$ (Bock & Aikin, 1981). Recalling that $\boldsymbol{\beta}_{jt} = \mathbf{C}_t\boldsymbol{\theta}_j$, $\bar{\boldsymbol{\beta}}_{jt}$ can be defined as

$$\bar{\boldsymbol{\beta}}_{jt} = \frac{\int_{\boldsymbol{\theta}} \mathbf{C}_t \boldsymbol{\theta} P(\mathbf{y}_j \mid \mathbf{w}_j, \mathbf{z}_j, \boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}}{P(\mathbf{y}_j \mid \mathbf{w}_j, \mathbf{z}_j)}.$$

As in the model estimation, Gauss-Hermite quadrature can be used to approximate the multidimensional integral. This yields the following approximation:

$$\bar{\boldsymbol{\beta}}_{jt} \approx \sum_{m=1}^{M} \mathbf{C}_t \boldsymbol{\theta}_m P\left(\boldsymbol{\theta}_m \mid \mathbf{y}_j \mathbf{w}_j, \mathbf{z}_j\right), \tag{8}$$

where the marginal posterior probabilities $P(\boldsymbol{\theta}_m \mid \mathbf{y}_j \mathbf{w}_j, \mathbf{z}_j)$ are obtained by Equation 10. Similarly, approximate posterior standard deviations can be obtained by

$$\operatorname{std}\left(\boldsymbol{\beta}_{jt}\right) \approx \sqrt{\sum_{m=1}^{M} \left(\mathbf{C}_t \boldsymbol{\theta}_m - \bar{\boldsymbol{\beta}}_{jt}\right)^2 P\left(\boldsymbol{\theta}_m \mid \mathbf{y}_j \mathbf{w}_j, \mathbf{z}_j\right)}.$$

These express the degree of reliability of the estimates of the random effects.

ANALYSIS OF THE TASK-VARIETY DATA

As mentioned earlier, the empirical application I use to illustrate the proposed method comes from a Dutch study on the effect of team characteristics on individual work conditions (Van Mierlo, 2003). A questionnaire was filled in by 886 employees from 88 teams of two organizations, a nursing home and a domiciliary care organization. Various aspects of the work condition were measured, one of which was the perceived task variety. The item wording of the five dichotomous items measuring perceived task variety is as follows (Van Veldhoven, De Jonge, Broerson, Kompier, & Meijman, 2002; translated from Dutch):

1. Do you always do the same things in your work?
2. Does your work require creativity?
3. Is your work diverse?
4. Does your work make enough usage of your skills and capacities?
5. Is there enough variation in your work?

The two answer categories are disagree and agree. Besides these five questionnaire items, we had information on four individual level covariates: year of birth (4 levels), number of years in the current job (3 levels), number of working hours per week (3 levels), and gender. A small portion of our sample (57 cases) had missing values on one or more of these items and covariates. These cases can, however, be retained in the analysis.

The data set on task variety was analyzed with the methods described using version 4.0 of the Latent GOLD program (Vermunt & Magidson, 2000). Missing data on the items was dealt within the ML estimation framework assuming that the missing data is missing at random (MAR). Whereas it is straightforward to obtain ML estimates of the parameters of a LC model with missing data on the item responses, missing data on covariates has to be dealt with in a more ad hoc manner. One option is, of course, to delete cases with missing values. Although the number of cases with one or more missing covariate values was rather small (35), I preferred to retain them in the analysis. For the four categorical covariates in the example, I assumed that the effect of the missing value category is equal to zero, Under effect or ANOVA-like coding, this amounts to assuming that a case with a missing value equals the average case as far as the covariate effect is concerned.

Table 1 reports the testing results—log-likelihood (LL), number of parameters, and Bayesian information criterion (BIC)—for the models that were estimated with the task-variety data. The models with random effects were estimated using 10 and 24 quadrature points, which in all cases yielded almost identical results. The BIC values for the standard one- to three-class models (Models A1, A2, and A3) show that a solution with two classes suffices. Model B2 is a two-class model without covariates but with a random intercept. From the comparison of the BIC values of Models A2 and B2, it can be seen that there is clear evidence for between-team variation in the latent distribution. This conclusion is not changed when including covariates in the model (compare Models C2 and D2). According to the BIC values, the models without covariates should be preferred over the models with covariates. The reason for this is that some of the fixed effects are not significant.

The model selection strategy I followed was to first determine the number of classes in a model without random effects nor covariates, and subsequently expand that model (in this case the two-class model) by including random effects and covariates. Though in most situations this will be a proper strategy, it cannot be ruled out that more latent classes are needed in the more extended models containing ran-

TABLE 1
Testing Results for the Estimated Models With the Task-Variety Data

Model	*Log-Likelihood*	*# Parameters*	*BIC Value*
A1. 1-class	–2797.0	5	5627.9
A2. 2-class	–2458.3	11	4991.2
A3. 3-class	–2443.9	17	5003.1
B2. 2-class with random intercept	–2436.0	12	4953.5
B3. 3-class with random intercept	–2418.4	19	4965.8
C2. 2-class with covariates	–2438.0	19	5005.0
C3. 3-class with covariates	–2417.2	33	5058.3
D2. 2-class with random intercept and covariates	–2420.4	20	4976.4
D3. 3-class with random intercept and covariates	–2388.1	35	5013.8

Note. BIC = Bayesian information criterion.

dom effects and/or predictors of class membership. That is the reason why I also estimated the three-class models B3, C3, and D3. Comparison of their BIC values with the ones of their two-class variants shows that there is no need to increase the number of classes from two to three according to this criterion. I also investigated whether inclusion of random effects for the four covariates (random slopes) improved model fit. The increase of the log-likelihood turned out to be very small (0.1, 1.9, 0.9, and 0.0, respectively), indicating that these random slopes are nonsignificant.

Table 2 reports the parameter estimates obtained with the estimated two-class models: Models A2, B2, C2, and D2. For identification, I fixed the parameters of

TABLE 2
Parameter Estimates Obtained With the Estimated Two-Class Models

	Model A2		*Model B2*		*Model C2*		*Model D2*	
Effect	*Est.*	*SE*	*Est.*	*SE*	*Est.*	*SE*	*Est.*	*SE*
MELR part								
Intercept	.61	.10	.76	.16	.86	.18	.90	.22
Year of Birth								
Before 1951					–.39	.17	–.34	.19
1951–1960					–.23	.13	–.17	.15
1961–1970					.13	.14	.10	.16
After 1970					.49	.17	.41	.20
Years in Current Job								
Less than 11					–.43	.14	–.33	.15
11–20					–.03	.15	–.11	.17
More than 20					.47	.20	.44	.22
Working Hours								
Less than 21					–.45	.12	–.52	.13
21–30					.33	.14	.34	.16
More than 30					.12	.14	.19	.16
Gender								
Male					–.09	.17	–.13	.20
Female					.09	.17	.13	.20
SD of the intercept			.96	.15			.91	.15
LC part								
Item 1, Class 1	.15	.02	.14	.02	.14	.02	.14	.02
Item 1, Class 2	.51	.02	.51	.02	.52	.02	.51	.02
Item 2, Class 1	.28	.03	.27	.03	.29	.03	.27	.03
Item 2, Class 2	.71	.02	.71	.02	.71	.02	.71	.02
Item 3, Class 1	.21	.04	.19	.03	.22	.04	.20	.03
Item 3, Class 2	.97	.01	.96	.01	.97	.01	.96	.01
Item 4, Class 1	.42	.03	.42	.03	.42	.03	.41	.03
Item 4, Class 2	.83	.02	.83	.02	.83	.02	.83	.02
Item 5, Class 1	.17	.03	.16	.03	.17	.03	.17	.03
Item 5, Class 2	.94	.02	.93	.02	.94	.02	.93	.02

Note. MELR = mixed-effects logistic regression; LC = latent class.

the first latent class to zero (dummy coding), and restricted the category-specific covariate effects to sum to zero (effect coding). Let us, however, first look at the parameters of the LC part of the model, which take on nearly the same values in all of the estimated two-class models. The reported conditional response probabilities describing the relationship between the latent variable and the five items correspond to the high task-variety response (disagree for item 1 and agree for the others). As can be seen, the first latent class can be labelled "low task-variety" and the second "high task-variety." Despite of the fact that there are only five items, the measurement part of the model is quite strong: the estimated proportion of classification errors is no more than 5% in each of the two-class models.

The estimate of the random effect in Model B2 shows that there is quite some between-team heterogeneity in the log odds of belonging to latent class two rather than to class one: the standard deviation of the random intercept equals .96. To get an impression of the meaning of this number on the probability scale, one can assign a value for θ_j in Equation 2 and substitute the obtained value for η_{ij2} in Equation 1. For example, with θ_j equal to –1.28 and 1.28—the z values corresponding to the lower and upper 10% tails of the normal distribution—we get latent class probabilities of .38 and .88, respectively. These numbers indicate that there is a quite large team effect on the perceived task variety: In the 10% "worst" teams, less than 38% of their employees perceived enough variation in the work, whereas this number is more than 88% in the 10% "best" teams. The intraclass correlation obtained with Equation 7 equals .22, which means that 22% of the total variance in perceived task variety can be explained by team membership.

The size of the standard deviation of the random intercept remained more or less the same when we included covariates in the model (compare Models B2 and D2), which indicates that team differences can not be explained by composition effects related to the four predictors. The covariate effects are similar in the models without (Model C2) and with (Model D2) a random intercept. The substantive interpretation of the fixed effects is as follows: The probability of belonging to the high task-variety class increases with Year of Birth (is higher for younger employees), increases with the number of Years in Current Job, is higher than average for persons working 21–30 hours per week and lower than average for the persons working less than 21 hours, and is slightly lower for males than for females.

Despite the fact that the substantive interpretation of the fixed effects are similar in Models C2 and D2, the significance tests for these effects are quite different. We determined the significance of the four predictor effects by means of multiparameter Wald tests. Table 3 reports the Wald values and their asymptotic p values obtained with Models C2 and D2. As can be seen, the Year of Birth and Years in Current Job effects that were significant at a 1% significance level in Model C2 were no longer significant when we included a random intercept. This illustrates the well-known phenomenon that p values may be underestimated when correlations between observations are not taken into account.

TABLE 3
Wald Statistics for Fixed Effects in Models C2 and D2

	Model C2			*Model D2*		
Covariate	*Wald*	*df*	*p*	*Wald*	*df*	*p*
Year of Birth	11.92	3	.01	6.19	3	.10
Years in Current Job	11.23	2	<.01	5.29	2	.07
Working Hours	15.62	2	<.01	15.23	2	<.01
Gender	0.29	1	.59	0.44	1	.51

As can be seen from the parameter estimates, the effects of Year of Birth and Years in Current Job are almost linear. With a more restricted specification in which these effects are assumed to be linear, the encountered effects are also significant—at a 5% significance level—in the model containing a random intercept (Wald = 5.95, $df = 1$, $p = .02$; and Wald = 4.89, $df = 1$, $p = .03$). However, again the values of the Wald statistics are much larger and thus seriously overestimated in the model without random effects (Wald = 11.49, $df = 1$, $p < .01$; and Wald = 10.33, $df = 1$, $p < .01$).

EXTENSIONS OF THE BASIC MODEL

Several of the extensions that have been proposed for the standard LC model could also be relevant in the context of the LC-MELR model. Three of these extensions are (a) models with items of other scale types, (b) models with ordered latent classes, and (c) models that relax the local independence assumption.

Suppose the items are not dichotomous or nominal but ordinal variables. The most natural way to use this information in the model specification is by restricting the relationship between the latent classes and the observed responses by means of an ordinal logit model (Agresti, 2002). One option is to use an adjacent-category logit ordinal model of the form

$$P\left(y_{ijk} = d_k \mid x_{ij} = t\right) = \frac{\exp\left(\delta_{d_k} + d_k \cdot \gamma_{kt}\right)}{\sum_{d_k=1}^{D_k} \exp\left(\delta_{d_k} + d_k \cdot \gamma_{kt}\right)},$$

which is similar to the nominal item response model but with the linear logistic restriction $\gamma_{d_k t} = d_k \cdot \gamma_{kt}$. Another option is to use a cumulative logit specification; that is,

$$P\left(y_{ijk} \geq d_k \mid x_{ij} = t\right) = \frac{\exp\left(\delta_{d_k} + \gamma_{kt}\right)}{1 + \exp\left(\delta_{d_k} + \gamma_{kt}\right)}.$$

Latent class models can be used not only with discrete items, but also with continuous items, yielding what is also referred to as a finite mixture model. For a continuous indicator, the conditional response probability $P(y_{ijk} = d_k \mid x_{ij} = t)$ is replaced by a conditional density having a certain parametric form, for example, a normal distribution with a class-specific mean and variance (Banfield & Raftery, 1993; Hunt & Jorgensen, 1999; Vermunt & Magidson, 2002).

A restriction that might be of interest in models with more than two latent classes is to assume that the classes are ordered. In the LC-MELR model, this involves constraining both the measurement model and the regression model for the latent classes. Suppose the items are ordinal and the relationship between latent classes and items is restricted with an adjacent category logit model. With ordered classes, this model has the following form:

$$P\left(y_{ijk} = d_k \mid x_{ij} = t\right) = \frac{\exp\left(\delta_{d_k} + d_k \cdot t \cdot \gamma_k\right)}{\sum_{d_k=1}^{D_k} \exp\left(\delta_{d_k} + d_k \cdot t \cdot \gamma_k\right)},$$

yielding what Heinen (1996) called a discretized IRT model and what Magidson and Vermunt (2001) called a LC factor model. The above model is, in fact, a discretized generalized partial credit model. For an extended discussion on the connection between restricted LC models and IRT models, see Vermunt (2001).

In order to complete the ordinal specification for the latent classes, the regression model for the latent variable should now be an ordinal instead of a nominal logistic regression model. An adjacent category ordinal logit model is obtained by restricting the term η_{ijt} appearing in Equation 1 to be equal to

$$\eta_{ijt} = \alpha_{0t} + \mathbf{w}_{ij}' \cdot t \cdot \boldsymbol{\alpha} + \mathbf{z}_{ij}' \cdot t \cdot \boldsymbol{\beta}_j .$$

Here, α_{0t} is a category-specific intercept, $\boldsymbol{\alpha}$ is a vector containing the other fixed effects, and $\boldsymbol{\beta}_j$ contains the random effects. The most important difference with the nominal model is that—except for the intercept—regression parameters are no longer category specific.

The last extension I would like to discuss is the possibility to relax the local independence assumption for particular item pairs (Hagenaars, 1988). This involves treating such a pair as a joint indicator variable. For example, if the first two indicators are locally dependent,

$$P\left(y_{ij1} = d_1, y_{ij2} = d_2 \mid x_{ij} = t\right) \neq P\left(y_{ij1} = d_1 \mid x_{ij} = t\right) P\left(y_{ij2} = d_2 \mid x_{ij} = t\right).$$

This means that we have to use the left-hand side instead of the right-hand side of this equation in the definition of the latent class model of interest. For $P(y_{ij1} = d_1, y_{ij2} = d_2 \mid x_{ij} = t)$, we will generally use a linear logistic model of the form

$$P\left(y_{ij1}=d_1, y_{ij2}=d_2 \mid x_{ij}=t\right)=\frac{\exp\left(\delta_{d_1}+\delta_{d_2}+\delta_{d_1d_2}+\gamma_{d_1t}+\gamma_{d_2t}\right)}{\sum_{d_1=1}^{D_1}\sum_{d_2=1}^{D_2}\exp\left(\delta_{d_1}+\delta_{d_2}+\delta_{d_1d_2}+\gamma_{d_1t}+\gamma_{d_2t}\right)},$$

in which the term $\delta_{d_1d_2}$ captures the within-class dependence between items 1 and 2.

DISCUSSION

An extension of the MELR model was proposed that can be used in situations in which the discrete dependent variable is a latent variable that is measured with multiple indicators. The proposed model can also be seen as an extension of the concomitant variable LC model to situations with clustered observations. To make ML estimation feasible, I adapted the E step of the EM algorithm making use of the conditional independence assumptions implied by the model. The new model was illustrated with an example from organizational research in which we constructed a latent task-variety typology. There was clear evidence for between-cluster variation in the latent class distribution, even after controlling for individual-level covariates.

A practical problem in the application of the LC-MELR model is that the numerical integration to be performed for parameter estimation can involve summation over a large number of points when the number of random effects is increased. Recall that the total number of quadrature points equals the product of the number of points used for each dimension. Despite the fact that the number of points per dimension may be somewhat reduced with multiple random effects, computational burden becomes enormous with more than 5 or 6 random coefficients. Adaptive quadrature may be a good alternative to the standard quadrature method used in this article. It has been shown that adaptive quadrature is quite accurate, even with a very small number of quadrature points (Rabe-Hesketh, Skrondal, & Pickles, 2002; Skrondal & Rabe-Hesketh, 2004). Other alternatives for solving the high-dimensional integrals are Bayesian MCMC and simulated likelihood methods, but these are very computationally intensive.

As shown by Vermunt and Van Dijk (2001), the practical problems with multiple random effects can be reduced by using a finite-mixture type random-effects model. In such a non-parametric ML approach, the distribution of the random effects is approximated with a small number of nodes whose locations and weights are unknown parameters to be estimated (Laird, 1978). This finite-mixture approach does not only have the advantage that it is computationally much less intensive, but also that it does not rely on unverifiable assumptions about the random effects (Aitkin, 1999). A recently simulation study by Agresti, Caffo, and Ohmand-Strickland (2004) showed that misspecifation of the random effects distribution may seriously affect efficiency of parameter estimates, and they therefore advocated the non-parametric approach. In agreement with their simulation results, using the finite-mixture approach in the task-variety application yielded similar parameter estimates but slightly larger values of the Wald statistics.

REFERENCES

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

Agresti A., Booth, J. G., Hobert, J. P., & Caffo, B. (2000). Random-effects modeling of categorical response data. *Sociological Methodology*, *30*, 27–80.

Agresti, A., Caffo, B., & Ohmand-Strickland, P. (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Computational Statistics & Data Analysis*, *47*, 639–653.

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, *55*, 218–234.

Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., & Rathouz, P. J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, *92*, 1375–1386.

Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics*, *49*, 803–821.

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, *41*, 164–171.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, *37*, 29–51.

Bock, R. D., & Aikin, M. (1981). Marginal maximum likelihood estimation of item parameters. *Psychometrika*, *46*, 443–459.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Application and data analysis*. Newbury Park, CA: Sage Publications.

Dayton, C. M., & Macready, G. B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, *83*, 173–178.

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

Goldstein, H. (1995) *Multilevel statistical models*. New York: Halsted Press.

Goldstein, H., & Browne, W. (2002). Multilevel factor analysis modelling using Markov chain Monte Carlo estimation. In G. A. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 225–243). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Hagenaars, J. A. (1988). Latent structure models with direct effects between indicators: Local dependence models. *Sociological Methods and Research*, *16*, 379–405.

Hagenaars, J. A. (1990). *Categorical longitudinal data — Loglinear analysis of panel, trend and cohort data*. Newbury Park, CA: Sage.

Hartzel, J., Agresti, A., & Caffo, B. (2001). Multinomial logit random effects models. *Statistical Modelling*, *1*, 81–102.

Hedeker, D. (1999). MIXNO: A computer program for mixed-effects nominal logistic regression. *Journal of Statistical Software*, *4*, 1–92.
Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. *Statistics in Medicine*, *22*, 1433–1446.
Hedeker, D., & Gibbons R. D. (1996). MIXOR: A computer program for mixed-effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine*, *49*, 157–176.
Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks, CA: Sage.
Hunt, L., & Jorgensen, M. (1999). Mixture model clustering using the MULTIMIX program. *Australian and New Zealand Journal of Statistics*, *41*, 153–172.
Juang, B. H., & Rabiner, L. R. (1991). Hidden Markov models for speech recognition. *Technometrics*, *33*, 251–272.
Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixture distribution. *Journal of the American Statistical Association*, *73*, 805–811.
Magidson, J., & Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays, *Sociological Methodology*, *31*, 223–264.
McCutcheon, A. L. (1987). *Latent class analysis*. Beverly Hills, CA: Sage.
Muthén, B., & Muthén, L. (2004). *Mplus 3.0: User's manual*. Los Angeles: Author.
Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2002). Reliable estimation of generalised linear mixed models using adaptive quadrature. *The Stata Journal*, *2*, 1–21.
Raudenbush, S. W., & Sampson, R. J. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology*, *29*, 1–41.
Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. London: Chapman & Hall/CRC.
Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis*. London: Sage.
Stroud, A. H., & Secrest, D. (1966). *Gaussian quadrature formulas*. Englewood Cliffs, NJ: Prentice Hall.
Van Mierlo, H. (2003). *Self-managing teams and psychological well-being*. Unpublished doctoral dissertation, Eindhoven University of Technology, The Netherlands.
Van Veldhoven, M., De Jonge, J., Broersen, S., Kompier, M., & Meijman, T. (2002). Specific relationships between psychosocial job conditions and job-related stress: A three-level analytic approach. *Work & Stress*, *16*, 207–228.
Vermunt, J. K. (2001). The use restricted latent class models for defining and testing nonparametric and parametric IRT models. *Applied Psychological Measurement*, *25*, 283–294.
Vermunt, J. K. (2002). An expectation-maximization algorithm for generalised linear three-level models. *Multilevel Modelling Newsletter*, *14*, 3–10.
Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology*, *33*, 213–239.
Vermunt, J. K., & Magidson, J. (2000). *Latent GOLD 2.0 user's guide.* Belmont, MA: Statistical Innovations, Inc.
Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge, MA: Cambridge University Press.
Vermunt, J. K., & Van Dijk, L. (2001). A nonparametric random-coefficients approach: The latent class regression model. *Multilevel Modelling Newsletter*, *13*, 6–13.
Wong, G. Y., & Mason, W. M. (1985). Hierarchical logistic models for multilevel analysis. *Journal of the American Statistical Association*, *80*, 513–524.

Accepted September 2004

APPENDIX

Implementation of the EM Algorithm

The preferred algorithm for obtaining ML estimates in LC analysis is the well-known EM algorithm. Contrary to Newton-Raphson and related methods, EM is a very stable algorithm that does not require good starting values (Hagenaars, 1990; Heinen, 1996). For EM, random starting values are good enough to converge to the nearest local maximum. When using the EM algorithm in the context of the LC-MELR model, one is, however, confronted with an important problem. The E step of the algorithm requires the computation of the joint conditional expectation of n_j latent class variables and R random effects for each level-2 unit j; that is, the posterior distribution $P(\mathbf{x}_j, \boldsymbol{\theta}_m \mid \mathbf{y}_j, \mathbf{w}_j, \mathbf{z}_j)$ containing $M \cdot T^{n_j}$ entries. This means that the ML estimation problem increases exponentially with the number of cases per group. With typical multilevel data group sizes, this may imply that one has to compute millions of probabilities, which is, of course, impractical. Fortunately, it turns out that the E step of EM can be adapted to the problem at hand.

Because of the structure of the LC-MELR model, the next step after obtaining the posterior probabilities $P(\mathbf{x}_j, \boldsymbol{\theta}_m \mid \mathbf{y}_j, \mathbf{w}_j, \mathbf{z}_j)$ would be to compute the marginal posterior probabilities for each level-2 unit, $P(x_{ij} = t, \boldsymbol{\theta}_m \mid \mathbf{y}_j, \mathbf{w}_j, \mathbf{z}_j)$ by collapsing over the latent class probabilities of the other level-1 units within level-2 unit j. In other words, in the E step we only need the n_j marginal posterior probability distributions containing $M \cdot T$ entries. This can be seen from the form of the (approximate) complete data log-likelihood, which is defined as

$$\log L_c = \sum_{m=1}^{M}\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{i=1}^{n_j} P\left(x_{ij} = t, \boldsymbol{\theta}_m \mid \mathbf{y}_j, \mathbf{w}_j, \mathbf{z}_j\right) \log P\left(x_{ij} = t, \mathbf{y}_{ij} \mid \mathbf{w}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\theta}_m\right). \quad (9)$$

It turns out that it is possible to compute $P(x_{ij} = t, \boldsymbol{\theta}_m \mid \mathbf{y}_j, \mathbf{w}_j, \mathbf{z}_j)$ without going through the full posterior distribution by making use of the conditional independence assumptions associated with the density function defined in Equation 5. In that sense, our procedure is similar to the forward-backward algorithm for the estimation of hidden Markov models for large numbers of time points (Baum, Petrie, Soules, & Weiss, 1970; Juang & Rabiner, 1991). A similar procedure was proposed by Vermunt (2002) for the estimation of generalized linear three-level models.

The marginal posterior probabilities $P(x_{ij} = t, \boldsymbol{\theta}_m \mid \mathbf{y}_j, \mathbf{w}_j, \mathbf{z}_j)$ can be decomposed as follows:

$$P\left(x_{ij} = t, \boldsymbol{\theta}_m \mid \mathbf{y}_j, \mathbf{w}_j, \mathbf{z}_j\right) = P\left(\boldsymbol{\theta}_m \mid \mathbf{y}_j, \mathbf{w}_j, \mathbf{z}_j\right) P\left(x_{ij} = t \mid \mathbf{y}_j, \mathbf{w}_j, \mathbf{z}_j, \boldsymbol{\theta}_m\right).$$

Our procedure makes use of the fact that in the LC-MELR

$$P\left(x_{ij}=t \mid \mathbf{y}_j,\mathbf{w}_j,\mathbf{z}_j,\boldsymbol{\theta}_m\right)=P\left(x_{ij}=t \mid \mathbf{y}_{ij},\mathbf{w}_{ij},\mathbf{z}_{ij},\boldsymbol{\theta}_m\right);$$

that is, conditionally on $\boldsymbol{\theta}$, x_{ij} is independent of the observed (and latent) variables of the other level-1 units within the same level-2 unit. This is the result of the fact that level-1 observations are mutually independent given the random effects, as is expressed in the density function described in Equation 5. Using this important result, we get the following slightly simplified decomposition:

$$P\left(x_{ij}=t,\boldsymbol{\theta}_m \mid \mathbf{y}_j,\mathbf{w}_j,\mathbf{z}_j\right)=P\left(\boldsymbol{\theta}_m \mid \mathbf{y}_j,\mathbf{w}_j,\mathbf{z}_j\right)P\left(x_{ij}=t \mid \mathbf{y}_{ij},\mathbf{w}_{ij},\mathbf{z}_{ij},\boldsymbol{\theta}_m\right).$$

The computation of the marginal posterior probabilities therefore reduces to the computation of the two terms on the right-hand side of this equation. The term $P(\boldsymbol{\theta}_m \mid \mathbf{y}_j, \mathbf{w}_j, \mathbf{z}_j)$ is obtained by

$$P\left(\boldsymbol{\theta}_m \mid \mathbf{y}_j,\mathbf{w}_j,\mathbf{z}_j\right)=\frac{P\left(\mathbf{y}_j,\boldsymbol{\theta}_m \mid \mathbf{w}_j,\mathbf{z}_j\right)}{P\left(\mathbf{y}_j \mid \mathbf{w}_j,\mathbf{z}_j\right)}, \tag{10}$$

where

$$P\left(\mathbf{y}_j,\boldsymbol{\theta}_m \mid \mathbf{w}_j,\mathbf{z}_j\right)=\pi\left(\boldsymbol{\theta}_m\right)\prod_{i=1}^{n_j}P\left(\mathbf{y}_{ij} \mid \mathbf{w}_{ij},\mathbf{z}_{ij},\boldsymbol{\theta}_m\right)$$
$$P\left(\mathbf{y}_j \mid \mathbf{w}_j,\mathbf{z}_j\right)=\sum_{m=1}^{M}P\left(\mathbf{y}_j,\boldsymbol{\theta}_m \mid \mathbf{w}_j,\mathbf{z}_j\right).$$

Notice that $P(\mathbf{y}_{ij} \mid \mathbf{w}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\theta}_m)$ was defined in Equation 4.

The other term, $P(x_{ij} = t \mid \mathbf{y}_{ij}, \mathbf{w}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\theta}_m)$, is computed by

$$P\left(x_{ij}=t \mid \mathbf{y}_{ij},\mathbf{w}_{ij},\mathbf{z}_{ij},\boldsymbol{\theta}_m\right)=\frac{P\left(x_{ij}=t,\mathbf{y}_{ij} \mid \mathbf{w}_{ij},\mathbf{z}_{ij},\boldsymbol{\theta}_m\right)}{P\left(\mathbf{y}_{ij} \mid \mathbf{w}_{ij},\mathbf{z}_{ij},\boldsymbol{\theta}_m\right)}.$$

As can be seen, the basic operation that has to be performed is the computation of $P(x_{ij} = t, \mathbf{y}_{ij} \mid \mathbf{w}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\theta}_m)$ for each i, j, t, and m by means of Equation 3. This shows that computation time increases only linearly with the number of level-1 observations instead of exponentially, as would be the case with a standard EM algorithm. Computation time can be reduced somewhat more by grouping records with the same values for the observed variables within level-2 units; that is, records with the values for $P(x_{ij} = t, \mathbf{y}_{ij} \mid \mathbf{w}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\theta}_m)$.

A practical problem in the above implementation of the ML estimation of the LC-MELR model is that underflows—numbers that cannot be distinguished from zero given the limited precision with which numbers can be stored—may occur in the computation of $P(\boldsymbol{\theta}_m \mid \mathbf{y}_j, \mathbf{w}_j, \mathbf{z}_j)$. More precisely, because it may involve multiplication of a large number ($R + n_j \cdot K$) of probabilities, the numerator of Equation 10 may become equal to zero for each m. Such underflows can, however, easily be prevented by working on a log scale. Letting

$$a_{jm} = \log\left[\pi(\boldsymbol{\theta}_m)\right] + \sum_i^{n_j} \log P\left(\mathbf{y}_{ij} \mid \mathbf{w}_{ij}, \mathbf{z}_{ij}, \boldsymbol{\theta}_m\right),$$

and $b_j = \max(a_{jm})$, $P(\boldsymbol{\theta}_m \mid \mathbf{y}_j, \mathbf{w}_j, \mathbf{z}_j)$ can be obtained by

$$P\left(\boldsymbol{\theta}_m \mid \mathbf{y}_j, \mathbf{w}_j, \mathbf{z}_j\right) = \frac{\exp\left(a_{jm} - b_j\right)}{\sum_r^M \exp\left(a_{jr} - b_j\right)}.$$

In the M step of the EM algorithm, the (approximate) complete data log-likelihood described in Equation 10 is improved by standard complete data algorithms for the ML estimation of multinomial logistic regression models.

Standard Errors and Identification

Contrary to Newton-like methods, the EM algorithm does not provide standard errors of the model parameters as a by-product. Estimated asymptotic standard errors can be obtained by computing the observed information matrix, which is the matrix of second-order derivatives of the log-likelihood function to all model parameters. The inverse of this matrix is the estimated variance-covariance matrix of the unknown model parameters.

The parameters of a LC model may not be identified when the number of items is too small given the postulated number of latent classes (Goodman, 1974). Nonidentification means that different sets of parameter values give the same maximum of the log-likelihood function. Identifiability can be checked by means of the information matrix. A sufficient condition for local identification is that all the eigenvalues of the information matrix are larger than zero, which yields a formal method to determine whether a model is identified. Similarly to standard LC analysis, it is not possible to provide a general rule for model identification of the LC-MELR model. It should, however, be noted that the dependence structure between observations belonging to the same cluster that is also used in the formulation of the log-likelihood function provides additional information on class membership of individuals. A sufficient condition for identification of the LC-MELR model is, therefore, that its standard LC variant without random effects is identified. This is, however, not a necessary condition: LC models that are not identified in their standard form may be identified in their mixed-effects form. For example, a two-class model with two nominal items is not identified, but its mixed effect variant is, even with only two individuals per group. With two individuals per group one has, in fact, a LC model with two correlated dichotomous latent variables and four items (two for each latent variable), and such a model is identified.

Software Implementation

The proposed LC-MELR model is implemented in version 4.0 of the Latent GOLD (Vermunt & Magidson, 2000). The implementation includes the extensions discussed earlier—ordinal and continuous items, ordinal latent variables, and local dependencies. The program also contains a variant of the LC-MELR model in which the random effects are assumed to come from an unspecified discrete distribution instead of a normal distribution (Vermunt, 2003). This yields what is sometimes referred to as a non-parametric random-effects model (Aitkin, 1999; Skrondal & Rabe-Hesketh, 2004).

To increase the likelihood of finding the global ML solution, Latent GOLD contains a search system that is based on using multiple sets of random starting values. According to my experience with the LC-MELR model, local maxima do not occur more often than in standard LC models. The exact algorithm implemented in Latent GOLD is a hybrid algorithm combining EM and Newton-Raphson; that is, EM iterations are performed till it is close enough to the maximum and then it switches to Newton-Raphson.

The LC-MELR model described in this article is also implemented in version 3.0 of the Mplus program (Muthén & Muthén, 2004). The manual, however, does not provide technical information on the estimation algorithm that is used to obtain the ML estimates of the model parameters.