

Latent Class Scaling Models for Longitudinal and Multilevel Data Sets

Jeroen K. Vermunt

Department of Methodology and Statistics

Tilburg University

Introduction

In the seventies various probabilistic generalizations of the well-known Guttman scaling method were proposed. One type of generalization yielded a class of latent class (LC) scaling models with names such as the Proctor (Proctor, 1970), latent distance (Lazarsfeld and Henry, 1968), and omission-inclusion (Dayton and Macready, 1976) model. A LC scaling model for J dichotomous items is a LC model with $J+1$ latent classes containing specific equality and inequality constraints on the class-specific response probabilities (Dayton, 1998; Heinen, 1996). These restrictions guarantee that classes are linearly ordered, which means that class k “scores” lower than class $k+1$. The beauty a LC scaling analysis is that, contrary to the more typical exploratory LC analysis, always well interpretable ordered classes are obtained. However, because the LC scaling assumptions do often not hold, also variants have been proposed that relax the linear order constraint (Dayton, 1998) or include unscalable classes (Goodman, 1975; Dayton & Macready, 1980).

Whereas the standard LC model (Goodman, 1974; Lazarsfeld and Henry, 1968; Dayton, 1998) was developed for the analysis of simple random sampling cross-sectional data, in the last decades various extensions have been proposed for dealing with data sets collected using more complex research designs. One of these is the latent Markov model, also referred to as latent transition or hidden Markov model, which can be used for the analysis of longitudinal data sets (Van de Pol & Langeheine, 1990; Collins & Wugalter, 1992; Vermunt, Tran, & Magidson, 2008). More specifically, it is a useful method when it makes sense to assume that individuals may move from one class to another across time points. More recently, Vermunt (2003) proposed an extension of the LC model for the analysis of multilevel

data sets. An example application could be the analysis of items measured on students which are nested within schools. This multilevel LC model accounts for the fact that the class memberships of lower-level units within the same higher-level unit are more likely to be the same than across different higher-level units. The model is somewhat similar to a LC model with concomitant variables (Dayton and Macready, 1988), but with the difference that the regression model for the latent classes contains not only fixed effects but also random effects (Vermunt, 2005). It is also similar to LC models for complex surveys data (Patterson, Dayton, and Graubard, 2002), but instead of applying a variance correction for the sampling design, the multilevel structure is made part of the model specification (Vermunt, 2002).

As far as I know, till now no connection has been made between LC scaling models and these more sophisticated LC models for longitudinal and multilevel data sets. The aim of this paper is to demonstrate how LC scaling models can be integrated into these recently developed extensions of LC analysis, which implies that LC scaling models are expanded by making them suitable for application with longitudinal and multilevel data sets. At the same time latent Markov and multilevel LC models are extended by allowing LC scaling restriction for the relationship between class membership and item responses. We will be able to investigate how groups differ in scaling class distribution and in predictor effects on the scaling classes, as well as how individuals change across scaling classes over time.

The remainder of this chapter is organized as follows. In the next three sections, I introduce the relevant LC scaling models, latent Markov models, and multilevel LC models, respectively. Then I present two LC scaling applications, one using a longitudinal data sets and one using a multilevel data set. The chapter ends with a few final remarks.

Latent Class Scaling Models

In this chapter we are dealing with LC models for dichotomous response variables, that is with false/true, incorrect/correct, disagree/agree, no/yes, or more generally 0/1 responses. I will denote the response of subject i on item j by y_{ij} – where $y_{ij} = 0, 1$ – and the number of items by J . The full response vector of a subject is denoted by \mathbf{y}_i . In addition to these J observed variables, a LC model contains a discrete latent variable. I will denote a subject's unobserved score on this latent variable by v_i , the number of LCs by C , and a particular class by c , where $c = 1, 2, \dots, C$.

LC analysis involves defining a model for $P(\mathbf{y}_i)$, the probability density of the multivariate response vector \mathbf{y}_i or, more specifically for the case of J dichotomous responses, the probability of answering the items according to one of the J^2 possible response patterns, for example, of answering the first two items correctly and the other ones incorrectly. The assumption underlying any type of LC or mixture model is that the density $P(\mathbf{y}_i)$ is a weighted average (or mixture) of the C class-specific densities $P(\mathbf{y}_i | \nu_i = c)$ (McLachlan & Peel, 2000). This is expressed mathematically as follows:

$$P(\mathbf{y}_i) = \sum_{c=1}^C P(\nu_i = c) P(\mathbf{y}_i | \nu_i = c). \quad (1)$$

The assumed mechanism by equation (1) is that each individual belongs to one of C exhaustive and mutually exclusive classes with probability $P(\nu_i = c)$ and that given membership of LC c one provides responses according to the probability density associated to this class. The classical LC model combines the assumption of equation (1) shared by all mixture models with the assumption of local independence (Dayton, 1998; Goodman, 1974). Local independence means that the J responses are mutually independent given a subject's class membership. It can be expressed as follows:

$$P(\mathbf{y}_i | \nu_i = c) = \prod_{j=1}^J P(y_{ij} | \nu_i = c). \quad (2)$$

Independence implies that the joint density $P(\mathbf{y}_i | \nu_i = c)$ is obtained as a product of the J item-specific densities $P(y_{ij} | \nu_i = c)$. Note that the local independence assumption is also used in other types of latent variables models, such as in factor analysis and IRT modeling, and is thus not specific for LC analysis. Combining the two basic equations (1) and (2) yields the following model for $P(\mathbf{y}_i)$:

$$P(\mathbf{y}_i) = \sum_{c=1}^C P(\nu_i = c) \prod_{j=1}^J P(y_{ij} | \nu_i = c). \quad (3)$$

To complete the model specification, we need to define the form of the conditional densities $P(y_{ij} | \nu_i = c)$. In a LC model for dichotomous items these are Bernoulli probability densities; that is,

$$P(y_{ij} | \nu_i = c) = \pi_{cj}^{y_{ij}} (1 - \pi_{cj})^{1 - y_{ij}},$$

Note this is a slightly complicated, but mathematically elegant, way to express that someone in LC c has a probability equal to $\pi_{cj} = P(y_{ij} = 1 | v_i = c)$ of giving response 1 to item j . Sometimes it is useful to parameterize the response probabilities π_{cj} using a logistic equation; that is, $\pi_{cj} = \frac{\exp(\beta_{cj})}{1 + \exp(\beta_{cj})}$ or $\beta_{cj} = \log \frac{\pi_{cj}}{1 - \pi_{cj}}$. Various types of constrained LC models have been proposed which involve imposing restrictions on the probabilities π_{cj} or their logits β_{cj} .

The constrained LC models I would like to discuss here are based on Guttman's (1947) notion of respondents and items being located along a common linear (ordinal) scale. Assume there are four 0/1 (negative/positive; incorrect/correct) items ($J=4$) and that these are ordered from most easy to most difficult. These items form a perfect Guttman scale if a positive response to the most difficult item 4 implies a positive response to the other items as well, a positive response to item 3 implies a positive response to items 1 and 2, and a positive response to item 2 implies a positive response to item 1. In other words, in a 4-item Guttman scale only the response patterns 0000, 1000, 1100, 1110, and 1111 may occur, which are the response patterns for the 5 ideal types. However, in practice, even with carefully constructed and administered items, other response patterns will be observed, for example, because of measurement errors. To tackle this problem, Proctor (1970) proposed a generalization of Guttman scaling incorporating response errors. This model is, in fact, a restricted LC model with $J+1$ latent classes, where each class corresponds to one of the ideal Guttman types and where θ_{cj} is the probability of giving a response which is inconsistent with the ideal type concerned; that is, $\theta_{cj} = \pi_{cj}$ for $c \leq j$ and $\theta_{cj} = 1 - \pi_{cj}$ for $c > j$. Dayton and Macready (1976) referred to these two types of errors as intrusions (when a 1 response replaces a 0) and omissions (when a 0 response replaces a 1).

Proctor's probabilistic version of the Guttman model assumes that the error probabilities are independent of the ideal type and the item; that is, $\theta_{cj} = \theta$. Dayton and Macready (1976) presented a modified version of Proctor's model with different probabilities for intrusions and omissions, implying that $\theta_{cj} = \theta_1$ for $c \leq j$ and $\theta_{cj} = \theta_2$ for $c > j$. Other response-error models have been proposed that assume that

error rates vary across items ($\theta_{cj} = \theta_j$) or across classes ($\theta_{cj} = \theta_c$). The latent distance model described by Lazarsfeld and Henry (1968) assumes that $\theta_{cj} = \theta_{j1}$ for $c \leq j$ and $\theta_{cj} = \theta_{j2}$ for $c > j$; that is, it has item-specific intrusion and omission probabilities.

Goodman (1975) proposed another type of probabilistic extension of the Guttman linear scales. Respondents from the $J+1$ ideal pattern latent classes respond exactly as expected ($\theta_{cj} = 0$) but one or more intrinsically unscalable classes are added with completely unrestricted response probabilities. Dayton and Macready (1980) generalized the Goodman model by allowing for errors in the latent classes corresponding to ideal patterns; that is, the response probabilities are unconstrained for the intrinsically unscalable class(es), but those for the ideal type classes are constrained to be in agreement with the Proctor model, the item-specific error model, etc..

Another extension of Guttman scaling involves the incorporation of more than a single linear scale underlying the observed responses (Dayton, 1998). For example, with four items, a combination of the linear Guttman scale orderings 0000, 0001, 0011, 0111, and 1111 and 0000, 0001, 0101, 0111, and 1111 yields a bifurm scale with six latent classes corresponding to the ideal patterns 0000, 0001, 0011, 0101, 0111, and 1111. In such a scale, items 2 and 3 do not show a prerequisite mutual relationship but a positive response to item 1 requires positive responses to items 2, 3, and 4, and a positive response to 2 and/or 3 requires a positive response to item 4. This can be extended to more complex multi-form scaling models, where errors can be taken into account.

Latent Markov Models

The latent Markov model – also referred to as hidden Markov model or latent transition model – is an extension of the LC model for the analysis of longitudinal data sets (Baum et al., 1970; Van de Pol & Langeheine, 1990; Collins & Wugalter, 1992; Vermunt, Tran, & Magidson, 2008). Similar as in standard LC models, individuals are assumed to belong to one of C latent classes at each of the measurement occasions, but what is different is that they are allowed to switch from one class to another across measurement occasions.

The notation should slightly be expanded to be able to deal with the longitudinal data structure. Individual i 's time-specific latent class membership,

response to item j , and vector of item responses will be denoted by v_{it} , y_{itj} , and \mathbf{y}_{it} , respectively, where $t = 0, 1, \dots, T_i$, with t denoting a particular time and T_i+1 being the total number of observations for individual i . A particular latent class at occasion t is denoted by c_t , where $c_t = 1, 2, \dots, C$. Using this notation, a latent Markov model can be defined as follows:

$$P(\mathbf{y}_i) = \sum_{c_0}^C \sum_{c_1}^C \dots \sum_{c_{T_i}=1}^C P(v_{i0} = c_0) \prod_{t=1}^{T_i} P(v_{it} = c_t | v_{i,t-1} = c_{t-1}) \prod_{t=0}^{T_i} P(\mathbf{y}_{it} | v_{it} = c_t)$$

where $P(v_{i0} = c_0)$ is the initial state probability, $P(v_{it} = c_t | v_{i,t-1} = c_{t-1})$ is the transition probability between adjacent time points, and $P(\mathbf{y}_{it} | v_{it} = c_t)$ is the time-specific conditional item density.

The specific form used for $P(\mathbf{y}_{it} | v_{it} = c_t)$ depends on the number of indicators and their scale types, where it should be noted that latent Markov models can be estimated with $J=1$, even when the single response variable is categorical. In this chapter, I will focus on latent Markov models for multiple dichotomous item responses per occasion. The new element is that the items will be assumed to form a probabilistic Guttman scale. More specifically, it is assumed that the J time-specific item responses are locally independence and Bernoulli distributed; i.e.,

$$P(\mathbf{y}_{it} | v_{it} = c) = \prod_{j=1}^J \pi_{cj}^{y_{itj}} (1 - \pi_{cj})^{1-y_{itj}},$$

and that, moreover, the class-specific response probabilities π_{cj} are in agreement with one of the probabilistic Guttman models discussed in the previous section. Note that it is also assumed that the measurement model is time-homogeneous, which can be seen from the fact that the π_{cj} do not contain an index t .

In latent Markov modeling, one will often also include explanatory variables affecting the initial state and the transition probabilities (Vermunt, Langeheine, and Böckenholt). Similar to LC modeling with concomitant variables, this involves defining logistic regression models for these model probabilities (Dayton and Macready, 1988). A regression model for the transition probabilities could be defined as follows:

$$\log \frac{P(v_{it} = c | v_{i,t-1} = d; \mathbf{z}_{it})}{P(v_{it} = c | v_{i,t-1} = c; \mathbf{z}_{it})} = \gamma_{0cd} + \sum_{r=1}^R \gamma_{rcd} z_{itr}, \text{ for } c \neq d. \quad (4)$$

Here z_{it} is a (time-varying) predictor and \mathbf{z}_{it} the vector of time-varying predictors. The term at the left-hand side is the logit of making a transition from state d to state c instead of staying in state c .

To apply the latent Markov model with more than a few time points, parameter estimation using maximum likelihood requires a special implementation of the E step of the EM algorithm, which is referred to as the Baum-Welch or forward-backward algorithm (Baum et al., 1970). This algorithm circumvents the computation and storage of the joint posterior distribution of the time-specific class memberships, which may have a very large number of entries, but instead computes and stores only the bivariate posterior distributions for adjacent time points. Vermunt, Tran, and Magidson (2008) showed how the forward-backward algorithm can be applied with a general class of latent Markov models, including models with multiple indicators, explanatory variables, multiple time-varying latent variables, mixture versions with one or more time-constant latent variable, and observations with missing values. This algorithm is implemented in the Latent GOLD software (Vermunt & Magidson, 2008).

Multilevel Latent Class analysis

Another common research design in applied fields such as social sciences, education, and health is the multilevel study design in which individuals are nested within groups. Analyses of such hierarchical data sets require the use of multilevel techniques. Recently, Vermunt (2003, 2005) proposed a multilevel extension of the LC model in which the class membership probabilities and/or the item response probabilities are allowed to vary randomly across groups. This model fits within a more general latent variable modeling framework described among others by Skrondal and Rabe-Hesketh (2004) and Vermunt (2008).

Let h denote a particular higher-level unit or group, H the number of groups, and n_h the number of individuals in group h . The index h will be used in v_{hi} , y_{hij} , and \mathbf{y}_{hi} , to indicate that it concerns a quantity of individual i belonging to group h . Moreover, \mathbf{y}_h is used to refer to the item responses of all members of higher-level unit h .

The simplest variant of the multilevel LC model – which is also the variant we will use here – is a model in which the class membership probabilities vary across

groups, but in which the item response probabilities are assumed to be the same across groups. In other words, observations within groups are dependent because they are more likely to belong to the same latent class. This model can be formulated as follows

$$P_h(\mathbf{y}_{hi}) = \sum_{c=1}^C P_h(v_{ih} = c) P(\mathbf{y}_{hi} | v_{hi} = c),$$

where the index h in $P_h(\cdot)$ indicates that the probably concerned is group specific, and where for our models $P(\mathbf{y}_{hi} | v_{hi} = c)$ has again the form

$$P(\mathbf{y}_{hi} | v_{hi} = c) = \prod_{j=1}^J \pi_{cj}^{y_{hij}} (1 - \pi_{cj})^{1 - y_{hij}},$$

and where, moreover, the class-specific response probabilities π_{cj} are in agreement with one of the probabilistic Guttman models discussed earlier. Note that the measurement model is assumed to be homogeneous across groups, which can be seen from the fact that the π_{cj} do not contain an index h .

In one variant of the multilevel LC model, Vermunt (2003, 2005) proposed modeling $P_h(v_{ih} = c)$ using a random effects logistic regression model; that is,

$$\log \frac{P_h(v_{ih} = c)}{P_h(v_{ih} = C)} = \alpha_c + \tau_c u_h, \text{ for } c < C,$$

where u_h is a normally distributed random effect with a mean equal to 0 and a variance equal to 1. When as in our applications the classes are ordered, the τ_c parameters can be further restricted as $\tau_c = c\tau$. This implies using an adjacent-category ordinal logit model for the class membership probabilities.

Similar to standard LC models, the multilevel LC model can be extended to include explanatory variables affecting the class memberships (Dayton & Macready, 1988). With ordinal classes, the adjacent-category logit model for the latent classes will have the following form:

$$\log \frac{P_h(v_{ih} = c + 1 | \mathbf{z}_{hi})}{P_h(v_{ih} = c | \mathbf{z}_{hi})} = \gamma_{0c} + \sum_{r=1}^R \gamma_{rc} z_{hir} + \tau u_h, \text{ for } c < C.$$

It is also possible to allow the random effects to depend of higher-level predictors, and to allow lower-level predictor effects on the class membership to vary across groups. In other words, a full multilevel logistic regression analysis can be performed for the class memberships (Vermunt, 2005).

As for the latent Markov models discussed above, a special variant of the E step of the EM algorithm is required for the maximum likelihood estimation of the multilevel LC model parameters. This algorithm which Vermunt (2003, 2008) called the upward-downward algorithm is implemented in the Latent GOLD software package (Vermunt & Magidson, 2003, 2008).

Application 1: Latent Markov scaling models for longitudinal data

The latent Markov models described above will be illustrated with the nine-wave National Youth Survey (Elliott, Huizinga, & Menard, 1989) for which data were collected annually from 1976 to 1980 and at three year intervals after 1980. At the first measurement occasion, the ages of the 1725 children varied between 11 and 17. To account for the unequal spacing across panel waves and to use age as the time scale, we define a model for 23 time points ($T + 1 = 23$), where $t = 0$ corresponds to age 11 and the last time point to age 33. For each subject, we have observed data for at most 9 time points (the average is 7.93) which means that the other time points are treated as missing values.

I will focus on the change in three dichotomous variables indicating whether young persons used alcohol, marijuana, and hard drugs (1=no; 2=yes) use during the past year. Scaling models with 4 latent states are specified, which differ in the assumed structure for the measurement error probabilities. The four ideal types correspond to non-users, alcohol users, alcohol and marijuana users, and users of all three types of drugs. To account for the fact that the transition probabilities may change over the 23 time periods (between age 11 and 33), I used age and age squared as time-varying predictors in the logit model for the transition probabilities.

{Insert Tables 1 about here}

Table 1 reports the log-likelihood values, the numbers of parameters, and the BIC and AIC values for the four estimated latent Markov scaling models. As can be seen, the Proctor model is clearly too restrictive. Both the model with item-specific errors and the one with separate intrusion and omission probabilities perform better

than the Proctor model. The latent distance model – which is the most general model with both features – is the preferred model for this data set.

{Insert Tables 2 about here }

Table 2 reports the class-specific response probabilities obtained with the Markov latent distance model. As can be seen the error probabilities are all rather small, with the largest been the omission probability for hard drugs. Note that while normally the latent distance model requires certain identifying restrictions, this is not the case when used in combination with a latent Markov model.

{Insert Table 3 about here }

Rather than reporting the initial state and the time-specific transition probabilities, I will show the estimated class proportions for each age. These are depicted in Figure 1. As can be seen, the size of the class of non-users decreases rapidly after age 11. The class of alcohol users increases monotonically from 2 percent at age 11 to 60 percent at age 30. Class 3 increases from 0 percent at age 11 to 29 percent at age 18 and decreases to 11 percent at age 33. The class of consumers all three types of drugs increases to 21 percent at age 22 and drops to 8 percent at age 33.

Application 2: Multilevel LC scaling models for nested data

This application uses a data set collected by Doolaard (1999), and which was also used by Fox and Glas (2001) to illustrate their multilevel IRT model. More specifically, information is available on a 18-item math test taken from 2156 pupils belonging to 97 schools in the Netherlands. The aim of the analysis is twofold: measuring pupils' math abilities and assessing differences between schools. For the first aim, we will use a LC scaling model for 9 of the 18 math items, while the second aim involves introducing school-level random coefficients in the LC scaling model. There is also information on individual-level covariates socioeconomic status (SES; standardized), non-verbal intelligence (ISI, standardized), and gender (0=males and 1=females), and a school-level covariate indicating whether a school participates in the national school leaving examination (CITO; 0=no, 1=yes).

{Insert Table 4 about here }

Table 4 reports the log-likelihood values, the numbers of parameters, and the BIC and AIC values for the estimated multilevel LC scaling models. As in the previous application, the latent distance model performs best because both item-specific errors and different intrusion and omission errors are needed. I also estimated 10-class LC models with logistic constraints yielding a Rasch and a two-parameter logistic model respectively. These two models perform somewhat better than the LC scaling models. This shows that assuming an S-shape relationship between the ordered classes and the responses seems to be better than a relationship in the form of a step function as assumed by the LC scaling models.

{Insert Tables 5 about here }

Table 5 presents the intrusion and omission probabilities obtained with the multilevel latent distance model. As can be seen, the intrusion probabilities are much larger than the ones we saw in the previous example, which indicates that the items are rather easy; that is, even the pupils with the lowest abilities have a rather high probability of making items 1 to 6 correctly.

The last model is the multilevel latent distance model with SES, ISI, gender and CITO predicting the class membership. This is specified using an ordinal logit model for the latent classes. Its much higher log-likelihood value and much lower BIC and AIC values compared to the model without covariates indicates that the covariates improve the model significantly.

{Insert 6 about here }

Table 6 reports the parameters of the (multilevel) ordinal logistic regression model for the latent classes. SES, ISI, and CITO have positive effects and gender has a negative effect. This shows that having a higher socioeconomic status, having a higher score on the nonverbal intelligence test, being a male, and belonging to a CITO school increases the likelihood of belonging to a higher ability class. It can also be seen that there remains a large amount of between-school variation after controlling for the four covariates. The size of the school effect can directly be compared with the

effects of SES and ISI because these are also standardized variables. The standard deviation of gender and CITO are about twice as small which means that we have to divide their effects by two if we want to compare these to other ones. The school effect turns out to be as important as the effect of a pupil's nonverbal intelligence (ISI) and more important than the other covariates. The differences between schools cannot be explained by differences in the composition of their populations.

Final remarks

This chapter demonstrated how LC scaling models can be used in longitudinal and multilevel studies. For this purpose, I proposed combining the probabilistic Guttman scaling models proposed by Proctor, Lazarsfeld, and Dayton and Macready with the more recently developed latent Markov and multilevel LC models. This new approach was illustrated using two empirical examples. The appendix describes the Latent GOLD syntax I used in these examples, so that interested readers will be able to apply these new models to their own data sets.

Appendix: Latent GOLD syntax for the two examples

As indicated in the text, the models presented in this chapter can be estimated with the Latent GOLD Syntax module. These are the “variables” and “equations” sections for the estimated Markov latent distance model:

```
variables
  caseid id;
  dependent alc, mrj, drugs;
  independent age11, age11_2;
  latent
    Class nominal dynamic 4;
equations
  Class[=0] <- 1;
  Class <- (~tra) 1 | Class[-1] + (~tra) age11 | Class[-1]
    + (~tra) age11_2 | Class[-1];
  alc <- (a1) 1 | Class;
  mrj <- (a2) 1 | Class;
  drugs <- (a3) 1 | Class;

  a1[2]=a1[4];a1[3]=a1[4]
  a2[2]=a2[1];a2[3]=a2[4]
  a3[2]=a3[1];a3[3]=a3[1]
```

The “variables” section defines the “caseid” connecting the multiple records of a person (each time point is a record in the data file), as well as defines the “dependent”, “independent”, and “latent” variables to be use in the model. In a latent Markov model, the “nominal” “latent” variable is defined to be “dynamic”.

The first two equations are the logit equations for the initial state and transition probabilities. Here, “1” denotes an intercept term. The specification of the second equation – with “(~tra)” and “| Class[-1]” – is such that it yields the parameterization described in equation (4) ; that is, coefficients which can be interpreted as effects on the logit corresponding to a certain transition.

The last three equations concern the three items. These imply that there are class-specific intercepts (logits of a positive response), which are labeled “a1”, “a2”, and “a3”, respectively. The restrictions defined for these logit parameters should be read as follows: for alcohol the logits for classes 2, 3, and 4 are equated, for marijuana the logits for classes 1 and 2 and of classes 3 and 4 are equated, and for hard drugs the logits for classes 1, 2, and 3 are equated.

The other estimated scaling models can be obtained with additional restrictions. The intrusion-omission model assumes that the errors are the same across items, which can be specified by inserting these two lines before the other constraints:

```
a2[1]=a1[1];a3[1]=a1[1];
a2[4]=a1[4];a3[4]=a1[4];
```

The item-specific error model assumes that the intrusion and omission probabilities are the same, or that the corresponding logits for positive responses have opposite signs. This can be specified by inserting this line before the other constraints:

```
a1[4]=-a1[1];a2[4]=-a2[1];a3[4]=-a3[1];
```

Combining the two sets of additional constraints yields the Proctor model; that is:

```
a2[1]=a1[1];a3[1]=a1[1];
a1[4]=-a1[1];a2[4]=-a2[1];a3[4]=-a3[1];
```

The definition of the restrictions on the item parameters is in the same way in multilevel LC models. The difference compared to the above latent Markov model occurs in the part concerning the latent variables. This is the syntax I used in the multilevel LC scaling application:

```
variables
  groupid school;
  independent cito, isi, ses, gender;
  dependent y1, y2, y3, y4, y5, y6, y7, y8, y9;
  latent
    Class ordinal 10,
    u continuous group;
equations
  Class <- 1 + cito + isi + ses + gender + u;
  y1 <- (a1) 1 | Class;
  y2 <- (a2) 1 | Class;
  y3 <- (a3) 1 | Class;
  y4 <- (a4) 1 | Class;
  y5 <- (a5) 1 | Class;
  y6 <- (a6) 1 | Class;
  y7 <- (a7) 1 | Class;
  y8 <- (a8) 1 | Class;
  y9 <- (a9) 1 | Class;

  restrictions on a1-a9 come here
```

As can be seen, a “groupid” connects the records of pupils belonging to the same school, and the group-level random effect “u” is defined as a “continuous” latent variable at the “group” level. The “equations” are the ordinal logit model for the latent classes (note that “Class” is defined to be “ordinal” instead of “nominal”) and the binary logit models for the items.

References

- 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.
- Collins, L. M., & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research*, *27*, 131-157.
- Dayton, C. M. (1998). *Latent class scaling models*. Thousand Oaks, CA: Sage
- Dayton, C. M. & Macready, G. B. (1976). A probabilistic model for validation of behavioral hierarchies. *Psychometrika*, *41*, 189-204.
- Dayton, C. M. & Macready, G. B. (1980). A scaling model with response errors and intrinsically unscalable respondents. *Psychometrika*, *45*, 343-356.
- Dayton, C. M., & Macready, G. B. (1988). Concomitant-variable latent class models. *Journal of the American Statistical Association*, *83*, 173-178.
- Doolaard, S. (1999). *Schools in Change or School in Chain*. Unpublished doctoral dissertation, University of Twente, The Netherlands.
- Elliott, D. S., Huizinga, D., & Menard, S. (1989). *Multiple problem youth: Delinquency, substance use, and mental health problems*. New York: Springer-Verlag.
- Fox, J. P., and Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, *66*, 269-286.
- Goodman, L. A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable: Part I - A modified latent structure approach. *American Journal of Sociology*, *79*, 1179-1259.
- Goodman, L. A. (1975). A new model for scaling response patterns: An application of the quasi-independence concept. *Journal of the American Statistical Association*, *70*, 755-768.
- Guttman, L. (1947). On Festinger's evaluation of scale analysis. *Psychological Bulletin*, *44*, 451-465.
- Hagenaars, J. A. (1990). *Categorical longitudinal data: Loglinear analysis of panel, trend and cohort data*. Newbury Park: Sage.
- Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks, CA: Sage.
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston, MA: Houghton Mifflin.

- McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- Patterson, B., Dayton, C. M. & Graubard, B. (2002). Latent class analysis of complex survey data: application to dietary data. *Journal of the American Statistical Association*, 97, 721-729.
- Poulsen, C. S. (1990). Mixed Markov and latent Markov modelling applied to brand choice behaviour. *International Journal of Research in Marketing*, 7, 519.
- Proctor, C. H. (1970). A probabilistic formulation and statistical analysis of Guttman scaling. *Psychometrika*, 35, 73-78.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. London: Chapman & Hall/CRC.
- Van de Pol, F., & Langeheine, R. (1990). Mixed Markov latent class models. *Sociological Methodology*, 213-247.
- Vermunt, J. K. (2001). The use of latent class models for defining and testing non-parametric and parametric item response theory models. *Applied Psychological Measurement*, 25, 283-294.
- Vermunt, J.K. (2002). Comments on “Latent class analysis of complex sample survey data”. *Journal of the American Statistical Association*, 97, 736-737.
- Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology*, 33, 213-239.
- Vermunt J.K (2005). Mixed-effects logistic regression models for indirectly observed outcome variables. *Multivariate Behavioral Research*, 40, 281-301.
- Vermunt, J.K. (2008). Multilevel latent variable modeling: An application in educational testing, *Austrian Journal of Statistics*, 37, 285–299.
- Vermunt, J. K., & Magidson, J. (2003). *Addendum to Latent GOLD User's Guide: Upgrade for Version 3.0*. Belmont, MA: Statistical Innovations Inc.
- Vermunt, J. K., & Magidson, J. (2008). *LG-syntax user's guide: Manual for Latent GOLD 4.5 Syntax Module*. Belmont, MA: Statistical Innovations.
- Vermunt, J. K. Langeheine, R., & Böckenholt, U. (1999). Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, 24, 178-205.

Vermunt, J. K., Tran, B., & Magidson, J. (2008). Latent class models in longitudinal research. In S. Menard (Ed.), *Handbook of longitudinal research: Design, measurement, and analysis* (pp. 373-385). Burlington, MA: Elsevier.

Table 1: Fit measures for the estimated latent Markov scaling models

Model	Log-likelihood	Number of parameters	BIC	AIC
Proctor	-13847	40	27993	27775
Intrusion-omission	-13840	41	27985	27762
Item-specific errors	-13827	42	27968	27739
Latent distance	-13720	45	27775	27530

Table 2: Response probabilities for the Markov latent distance model

		Latent scale type			
		c=1	c=2	c=3	c=4
Alcohol	no	0.90	0.02	0.02	0.02
	yes	0.10	0.98	0.98	0.98
Marijuana	no	0.98	0.98	0.09	0.09
	yes	0.02	0.02	0.91	0.91
Hard drugs	no	0.99	0.99	0.99	0.15
	yes	0.01	0.01	0.01	0.85

Table 3: Age-specific latent scales type proportions for the Markov latent distance model

Age	Latent scale type			
	c=1	c=2	c=3	c=4
11	0.98	0.02	0.00	0.00
12	0.83	0.15	0.02	0.00
13	0.67	0.24	0.07	0.02
14	0.51	0.30	0.14	0.05
15	0.37	0.33	0.21	0.09
16	0.26	0.35	0.26	0.14
17	0.19	0.35	0.28	0.17
18	0.15	0.36	0.29	0.19
19	0.13	0.37	0.29	0.21
20	0.11	0.39	0.28	0.22
21	0.10	0.41	0.26	0.22
22	0.10	0.44	0.24	0.22
23	0.10	0.47	0.22	0.21
24	0.11	0.49	0.20	0.20
25	0.12	0.52	0.18	0.19
26	0.12	0.54	0.16	0.17
27	0.13	0.56	0.15	0.15
28	0.14	0.58	0.14	0.14
29	0.15	0.59	0.13	0.12
30	0.17	0.60	0.12	0.11
31	0.18	0.60	0.12	0.10
32	0.19	0.60	0.12	0.09
33	0.21	0.59	0.11	0.08

Table 4: Fit measures for the estimated multilevel latent class scaling models

Model	Log-likelihood	Number of parameters	BIC	AIC
Proctor	-10496	11	21076	21014
Intrusion-omission	-10203	12	20498	20430
Item-specific errors	-10274	19	20693	20585
Latent distance	-10077	28	20369	20210
Rasch	-10068	20	20289	20175
Two-parameter logistic	-10032	28	20279	20120
Latent distance + covariates	-9850	32	19946	19764

Table 5: Intrusion and omission probabilities for the estimated multilevel latent distance model

Item	Intrusion	Omission
1	0.44	0.02
2	0.63	0.04
3	0.47	0.06
4	0.50	0.06
5	0.53	0.15
6	0.43	0.15
7	0.33	0.12
8	0.30	0.19
9	0.30	0.25

Table 6: Parameters of the ordinal logistic model for the latent classes obtained with latent distance model with covariates

	coefficient	S.E.	Wald	DF	p-value
School (random)	0.195	0.019	100.18	1	1.40E-23
CITO	0.178	0.048	13.84	1	0.0002
ISI	0.180	0.015	139.38	1	3.60E-32
SES	0.113	0.014	68.80	1	1.10E-16
Gender	-0.086	0.021	16.70	1	4.40E-05