Measurement and Research Department Reports 97-8

Estimating Multiple Classification Latent Class Models

E. Maris

Cito

Measurement and Research Department Reports 97-8

Estimating Multiple Classification Latent Class Models

E. Maris

Cito
Arnhem, 1997

Abstract

This paper presents a new class of models for persons-by-items data. The essential new feature of this class is the representation of the persons: every person is represented by its membership to *multiple latent classes*, each of which belongs to one *latent classification*. The models can be considered as a formalization of the hypothesis that the responses come about in a process that involves the application of a number of *mental operations*. Two algorithms for maximum likelihood (ML) and maximum a posteriori (MAP) estimation are described. They both make use of the tractability of the complete data likelihood to maximize the observed data likelihood. Properties of the MAP estimators (i.e., uniqueness and goodness-of-recovery) and the existence of asymptotic standard errors were examined in a simulation study. Then, one of these models is applied to the responses to a set of fraction addition problems. Finally, the models are compared to some related models in the literature.

Key-words: latent class models, latent response models, cognitive processes, EM-algorithm.

This paper presents a new class of models for persons-by-items data. The essential new feature of this class is the representation of the persons: every person is represented by its membership to *multiple latent classes*.

In traditional latent class analysis (Lazarsfeld & Henry, 1968; Goodman, 1974) every person is represented by its membership to one of T latent classes. In this paper, such a set of T latent classes is called a *latent classification*. The models to be presented here involve more than one latent classification, and every person is thus characterized by its memberships to the latent classes of these *multiple* latent classifications. Therefore, they are called *multiple classification latent class models* (MCLCM's).

MCLCM's can be considered as a formalization of the hypothesis that the responses come about in a process that involves the application of a number of *mental operations*. Each of these mental operations corresponds to one latent classification. With binary latent classifications, one of the classes in every classification corresponds to mastery of this mental operation and the other to non-mastery.

We start by giving an example of an item type (simplifying fractions) whose solution process can be described in terms of two mental processes (splitting and identifying).

1 Example: Simplifying Fractions

Consider items of the following type: $9/4 = \ldots$, $3/2 = \ldots$, $2/6 = \ldots$, and $14/5 = \ldots$. The subject's task is to simplify these fractions as much as possible. For example, the correct answer to $14/5 = \ldots$ is $2\ \ 4/5$. It is assumed that, to be able to correctly simplify all possible fractions, one has to master two mental operations: *splitting* and *identifying*. The splitting operation involves that a given fraction is split in a units part and a fraction part. And the identifying operation involves that the largest common denominator of the fraction part is identifiable.

Given these two mental operations, four types of items can be distinguished. First, there are items that require neither splitting nor identifying. An example is $3/5 = \ldots$. The correct answer is simply $3/5$. No mental operation has to be performed to produce this answer. Second, there are items that require only identifying. An example is $2/6 = \ldots$. The subject has to find the largest common denominator (i.e., 2) and then divide numerator and denominator by it. This produces $1/3$ as the correct answer. Third, there are items that require only splitting. An example is $5/3$. Applying the splitting operation, $5/3$ is transformed into $3/3 + 2/3$. Then, $3/3$ is transformed into 1 without having to identify the largest common denominator. And fourth, there are items that require both splitting and identifying. An example is $8/6 = \ldots$. Applying the splitting operation, $8/6$ is transformed into $6/6 + 2/6 = 1 + 2/6$. And applying the identifying operation, the largest common denominator of 2 and 6 is seen to be 2 such that $2/6$ can be written as $1/3$.

Analogous to the distinction of four types of items, also four types of persons are distinguished: those that master neither splitting nor identifying, those that master only splitting or identifying, and those that master both splitting and identifying. For each of these four types of persons, one can specify which item types they will answer correctly. For this, the rule has to be followed that an item is answered correctly if and only if the person masters all mental operations required by the item. This is illustrated in Figure 1. In this figure, a 1 is used to denote that an item requires a particular mental operation and 0 to denote that it does not. Similarly for the persons, a 1 is used to denote that a person masters a particular mental operation and 0 to denote that he or she does not.

Splitting		Requires	0	1	0	1
	Identifying		0	0	1	1
Masters						
0	0		1	0	0	0
1	0		1	1	0	0
0	1		1	0	1	0
1	1		1	1	1	1

Figure 1
Responses of four types of persons to four types of items as a function of the mental operations (splitting and identifying) mastered by the persons and required by the items

2 Models

First, we present the *conjunctive MCLCM*. Then, other MCLCM's are presented along the same lines.

2.1 The Conjunctive MCLCM

MCLCM's consist of two components: (a) a model for the latent class memberships and (b) a model for the item responses conditional on the latent class memberships. Although our main interest is in (b) (the way the item responses come about), the exposition is facilitated by considering (a) first.

2.1.1 The saturated model for the latent class memberships

We consider the random variable (RV) Z_k that indicates a person's membership to some class of the k-th latent classification. The number of latent classifications is denoted by K ($k = 1, \ldots, K$). In general, Z_k is a categorical RV with a number of categories that can be specified freely, as in ordinary latent class analysis. In this paper, we only consider latent classifications consisting of *two* classes. Thus, Z_k is a Bernoulli RV. The realizations of Z_k are denoted by z_k whose values are 0 or 1. The restriction to two classes is motivated by the fact that, in this way, the latent classifications can be interpreted as mental operations, with the two classes corresponding to mastery and non-mastery.

Because there are K latent classifications, every person is characterized by a K-dimensional vector of z_k's. This vector is denoted by $\boldsymbol{z}$ ($\boldsymbol{z}=(z_1, \ldots, z_K)^t$). This vector $\boldsymbol{z}$ is a realization of the K-dimensional RV $\boldsymbol{Z}$ ($\boldsymbol{Z}=(Z_1, \ldots, Z_K)^t$). For $\boldsymbol{Z}$, several models can be formulated. A model for $\boldsymbol{Z}$ is a probability density function (PDF). This PDF depends on some vector of parameters $\boldsymbol{\xi}$, and is denoted by $P(\boldsymbol{Z} = \boldsymbol{z}; \boldsymbol{\xi})$.

We consider the saturated model for $\boldsymbol{Z}$. This model assumes that the probabilities of each of the different realizations of Z ($\boldsymbol{z}$-patterns) are unrestricted. allowing for all possible statistical dependencies between the Z_k's. The total number of $\boldsymbol{z}$-patterns is 2^K. This number is denoted by T, and t is used as an index for the $\boldsymbol{z}$-patterns ($t = 1, \ldots, T$). The only restrictions on the T parameters $\xi_t := P(\boldsymbol{Z} = \boldsymbol{z}_t)$ are the obvious ones, $0 < \xi_t < 1$ and the fact that they have to sum to 1. This model is characterized by the following PDF:

$$P(\boldsymbol{Z} = \boldsymbol{z}; \boldsymbol{\xi}) = \prod_{t=1}^{T} \xi_t^{I_t(\boldsymbol{z})} \qquad (1)$$

in which I_t is an indicator function having the value 1 if $\boldsymbol{z}$ is the t-th pattern and 0 otherwise.

2.1.2 A conjunctive latent response model for the item responses conditional on Z

Before describing the conjunctive *latent response model* (LRM), some notation and assumptions are presented that are common to all models for the item responses.

Notation and general assumptions. For every person, a vector of I item responses is observed. The RV Y_i is used to denote the response on the i-th item ($i = 1, \ldots, I$). In general, Y_i is a categorical RV with any number of categories, but in this paper only dichotomous items (e.g., correct/incorrect) are considered. The realizations of Y_i are denoted by y_i whose values are either 0 or 1. The vector-valued RV of the I item responses is denoted by $\boldsymbol{Y}$ $(= (Y_1, \ldots, Y_I)^t)$ and its realization by $\boldsymbol{y}$ $(= (y_1, \ldots, y_I)^t)$.

The PDF of $\boldsymbol{Y}$ depends on $\boldsymbol{Z}$, the vector of latent class memberships, and $\boldsymbol{\eta}$, a vector of item parameters. This PDF is denoted by $P(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{Z} = \boldsymbol{z}; \boldsymbol{\eta})$ or $P(\boldsymbol{Y}|\boldsymbol{Z}; \boldsymbol{\eta})$ if there is no danger of confusion. The assumption of *local statistical independence* (LSI) is made, involving that, conditionally on $\boldsymbol{Z}$, the Y_i's are statistically independent. Denoting the PDF of Y_i by $P(Y_i|\boldsymbol{Z}; \boldsymbol{\eta})$, this assumption can be expressed as follows:

$$P(\boldsymbol{Y}|\boldsymbol{Z}; \boldsymbol{\eta}) = \prod_{i=1}^{I} P(Y_i|\boldsymbol{Z}; \boldsymbol{\eta}). \tag{2}$$

It is also assumed that $P(Y_i|\boldsymbol{Z}; \boldsymbol{\eta})$ depends on $\boldsymbol{\eta}$ only through some subset $\boldsymbol{\eta}_i$ of $\boldsymbol{\eta}$, the item parameters of the i-th item. Therefore, $P(Y_i|\boldsymbol{Z}; \boldsymbol{\eta})$ can also be expressed as $P(Y_i|\boldsymbol{Z}; \boldsymbol{\eta}_i)$. We now consider a further specification of $P(Y_i|\boldsymbol{Z}; \boldsymbol{\eta}_i)$, namely as a LRM.

A conjunctive latent response model. The basic idea behind LRM's is that the observed response (Y_i in our case) is the result of a mapping that takes a set of *latent responses* as its argument (Maris, 1995; Maris, De Boeck and Van Mechelen, 1996). LRM's are defined by (a) a model for the latent responses, and (b) a so-called *condensation rule* that maps these latent responses into the observed responses. We first consider (a). Considering the latent classifications as corresponding to mental operations (e.g., splitting and identifying), we now consider the K latent responses that are the results of the application of each of these K mental operations. These latent responses denote the successful or non-successful application of these mental operations. Thus, the latent responses are dichotomous. When speaking about latent responses, the term *component* will be used to denote the process that generates the latent response. Every component corresponds to one latent classification.

The k-th latent response on item i is denoted by the RV X_{ik}. In general, X_{ik} is a categorical RV that can have any number of categories, but here only dichotomous X_{ik}'s are considered. For example, simplifying a fraction may involve X_{i1} and X_{i2}, denoting, respectively, whether the splitting and the identifying operation was applied successfully ($X_{i1}, X_{i2} = 1$) or not ($X_{i1}, X_{i2} = 0$). The vector-valued RV of the K latent responses on item i is denoted by $\boldsymbol{X}_i$ $(= (X_{i1}, \ldots, X_{iK})^t)$. The PDF of $\boldsymbol{X}_i$ is denoted by $P(\boldsymbol{X}_i|\boldsymbol{Z};\boldsymbol{\eta}_i)$. We make the assumption of LSI, which involves that, conditionally on $\boldsymbol{Z}$, the X_{ik}'s are statistically independent. Denoting the PDF of X_{ik} by $P(X_{ik}|\boldsymbol{Z};\boldsymbol{\eta}_i)$, this assumption is expressed as follows:

$$P(\boldsymbol{X}_i|\boldsymbol{Z};\boldsymbol{\eta}_i) = \prod_{k=1}^{K} P(X_{ik}|\boldsymbol{Z};\boldsymbol{\eta}_i) \tag{3}$$

(For situations in which this assumption is violated, the models can be extended in a straightforward way.)

Next, it is assumed that $P(X_{ik}|\boldsymbol{Z};\boldsymbol{\eta}_i)$ depends on $\boldsymbol{\eta}_i$ only through some subset $\boldsymbol{\eta}_{ik}$ of $\boldsymbol{\eta}_i$. Therefore, $P(X_{ik}|\boldsymbol{Z};\boldsymbol{\eta}_i)$ can also be expressed as $P(X_{ik}|\boldsymbol{Z};\boldsymbol{\eta}_{ik})$. This subset of the item parameters will be called the *component item parameters.*

Finally, it is assumed that $P(X_{ik}|\boldsymbol{Z};\boldsymbol{\eta}_{ik})$ depends on $\boldsymbol{Z}$ only through Z_k, the membership in the latent classification that corresponds to this component. It follows that $P(X_{ik}|\boldsymbol{Z};\boldsymbol{\eta}_{ik})$ can be replaced by $P(X_{ik}|Z_k;\boldsymbol{\eta}_{ik})$. Considering the memberships in the latent classifications as mastery or non-mastery of a mental operation, this assumption involves that the result of the application of the k-th mental operation (i.e., X_{ik}) only depends on the person's mastery of this particular mental operation. This assumption is necessary for the substantive interpretation of the Z_k's. In particular, if the parameter estimates show that some component is only involved in a particular item type (e.g., items that involve splitting), then we know that the corresponding Z_k operates on the item characteristic that distinguishes this item type from the others.

We now consider the different component item parameters. For every (item,component)-pair there are two parameters. Every parameter corresponds to one conditional probability: $P(X_{ik} = 1|Z_k = 1)$, the probability of a correct application of the k-th mental operation given that it is mastered, and $P(X_{ik} = 1|Z_k = 0)$, the probability of a correct application of this mental operation given that it is *not* mastered. $P(X_{ik} = 1|Z_k = 0)$ is the probability that non-mastery of the k-th mental operation can be compensated by other mental resources. And *one minus* $P(X_{ik} = 1|Z_k = 1)$ is the probability of a careless error (e.g., due to lack-of-attention). The conditional probabilities $P(X_{ik} = 1|Z_k = 1)$ and $P(X_{ik} = 1|Z_k = 0)$ are the item parameters of this model and they will be denoted by, respectively, η_{ik1} and η_{ik0}. Thus, $\boldsymbol{\eta}_{ik} = (\eta_{ik1}, \eta_{ik0})^t$. This allows us to write $P(X_{ik}|Z_k; \boldsymbol{\eta}_{ik})$ as follows:

$$P(X_{ik}|Z_k; \boldsymbol{\eta}_{ik}) = \left(\eta_{ik1}^{X_{ik}}(1 - \eta_{ik1})^{1-X_{ik}}\right)^{Z_k} \times \left(\eta_{ik0}^{X_{ik}}(1 - \eta_{ik0})^{1-X_{ik}}\right)^{1-Z_k} \tag{4}$$

The second defining characteristic of a LRM is its condensation rule. This condensation rule is a mapping of $\boldsymbol{X}_i$ into Y_i, denoted by $\mathcal{C}(\boldsymbol{X}_i)$. The *conjunctive* condensation rule is defined as follows:

$$\mathcal{C}(\boldsymbol{X}_i) = \prod_{k=1}^{K} X_{ik}$$

This function has the value 1 if and only if all X_{ik}'s have the value 1. A useful interpretation of this condensation rule is in terms of mental operations whose successful application is *necessary* for giving the correct response. Thus, *all* mental operations have to be applied successfully to give a correct response. For example, to solve a fraction item, both the splitting and the identifying operation have to be applied successfully (if they are involved in the solution process, of course). Together with the assumption of LSI of the X_{ik}'s (see (3)), this condensation rule leads to the following form for $P(Y_i|\boldsymbol{Z}; \boldsymbol{\eta}_i)$:

$$P(Y_i|\boldsymbol{Z}; \boldsymbol{\eta}_i) = \left(\prod_{k=1}^{K} P(X_{ik} = 1|Z_k)\right)^{Y_i} \left(1 - \prod_{k=1}^{K} P(X_{ik} = 1|Z_k)\right)^{1-Y_i} \tag{5}$$

2.1.3 The marginal model for the item responses

Finally, we consider the PDF of $\boldsymbol{Y}$, the vector of observed item responses. The starting-point for this PDF is the *joint* PDF of $\boldsymbol{Y}$ and $\boldsymbol{Z}$:

$$P(\boldsymbol{Y}, \boldsymbol{Z}; \boldsymbol{\eta}, \boldsymbol{\xi}) = P(\boldsymbol{Y}|\boldsymbol{Z}; \boldsymbol{\eta})P(\boldsymbol{Z}; \boldsymbol{\xi})$$

From this joint PDF we get the marginal PDF of $\boldsymbol{Y}$ by summing over all possible realizations of $\boldsymbol{Z}$:

$$P(\boldsymbol{Y}; \boldsymbol{\eta}, \boldsymbol{\xi}) = \sum_{z_1=0}^{1} \cdots \sum_{z_K=0}^{1} P(\boldsymbol{Y}, \boldsymbol{Z} = \boldsymbol{z}; \boldsymbol{\eta}, \boldsymbol{\xi})$$

2.2 Other MCLCM's

Other MCLCM's are obtained by formulating (a) a different model for the latent class memberships, and (b) a different model for the item responses conditional on $\boldsymbol{Z}$.

2.2.1 Different models for the latent class memberships

Another model for $\boldsymbol{Z}$ is the independence model. This model assumes that every Z_k is independently distributed. Because Z_k is a Bernoulli PDF, it is characterized by a single parameter ξ_k which is equal to $P(Z_k = 1)$. The only restriction on these ξ_k's is that they have to be between 0 and 1. This model is characterized by the following PDF:

$$P(\boldsymbol{Z} = \boldsymbol{z}; \boldsymbol{\xi}) = \prod_{k=1}^{K} \xi_k^{z_k}(1 - \xi_k)^{1-z_k} \tag{6}$$

Still another model for $\boldsymbol{Z}$ is a loglinear model for the K-dimensional $2 \times 2 \times \ldots \times 2$-table of latent class memberships. Actually, the models in (1) and (6) can also be formulated as loglinear models for this table: a model with all possible interactions between the K classifications for (1) and a model with only main effects for (6). Obviously, by adding and deleting interaction terms a whole variety of loglinear models can be formulated.

2.2.2 Different models for the item responses conditional on Z

Different latent response models. LRM's are very well suited for formalizing psychological hypotheses about the process in which the responses come about. Within this framework, different types of MCLCM's, corresponding to different psychological hypotheses, can be formulated easily. Three kinds of extensions to the conjunctive model are possible. First, one can formulate different PDF's for the latent responses $\boldsymbol{X}_i$ conditional on $\boldsymbol{Z}$. For example, this PDF may involve statistical dependencies between the latent responses, reflecting the hypothesis that the application of one mental operation influences the application of others. Second, one can assume the latent responses to be polytomous, or even continuous, instead of dichotomous. For example, continuous latent responses may be an appropriate choice for a detection or identifiability task in which stimuli are presented that vary on a number of continuous dimensions. And third, one can formulate other condensation rules besides the conjunctive one.

Another useful condensation rule is the *disjunctive* one. It is defined as follows:

$$\mathcal{C}(\boldsymbol{X}_i) = 1 - \prod_{k=1}^{K}(1 - X_{ik})$$

This function has the value 1 if and only if there is at least one X_{ik} that has the value 1. A useful interpretation of this condensation rule is in terms of mental operations or *strategies* whose successful application is *sufficient* for giving a correct response. Together with the assumption of LSI of the X_{ik}'s, this condensation rule leads to the following form for $P(Y_i|\boldsymbol{Z};\boldsymbol{\eta}_i)$:

$$P(Y_i|\boldsymbol{Z};\boldsymbol{\eta}_i) = \left(1 - \prod_{k=1}^{K} P(X_{ik} = 0|Z_k)\right)^{Y_i} \left(\prod_{k=1}^{K} P(X_{ik} = 0|Z_k)\right)^{1-Y_i} \qquad (7)$$

Still other condensation rules may have more than two different function values. For example, one can formulate MCLCM's for multiple choice items by chosing the condensation rule such that every pattern of latent responses is mapped into a particular response alternative according to some hypothesis about the response process. Such an hypothesis should not only specify how the correct response comes about, but also the different *incorrect* responses.

Estimation (Section 4) will be considered only for the conjunctive and the disjunctive model with independent latent responses. The extension to other condensation rules and other models for the latent responses is straightforward, however.

Restrictions on the item parameters. Besides extending the LRM-framework by formulating other PDF's for the latent responses and using other condensation rules, the usefulness of this class of models is also enhanced by introducing restrictions on the item parameters. In particular, interesting special cases appear if η_{ik0} and/or η_{ik1} are fixed at 0 or 1. Under the conjunctive condensation rule, fixing η_{ik0} at 0, the restriction is imposed that this item absolutely requires mastery of this mental operation. This type of restrictions is very well suited for testing hypotheses about the response process. For example, one can fix η_{i10} at 0 for all fraction items that require splitting, and fix η_{i20} at 0 for all fraction items that require identifying. The introductory example was implicitely based on this kind of deterministic response model (see Figure 1). Also under the conjunctive condensation rule, by fixing *both* η_{ik0} and η_{ik1} at 1, the restriction is imposed that the corresponding mental operation is simply not involved in the solution of this item. For example, the fraction $2/6 = \ldots$ does not involve splitting. So, one can fix η_{i10} and η_{i11} at 1 for this item. Under the disjunctive condensation rule, similar restrictions can be imposed.

Latent response and latent variable models. At this point, we should point out the difference between LRM's and latent variable models in general. In a broad sense, LRM's *are* latent variable models because the model for the observed data (the Y_i's) is obtained by integrating (summing) out a set of unobserved random variables (the $\boldsymbol{Z}$'s and the $\boldsymbol{X}_i$'s). In a narrow sense, latent variable models (the factor analysis model, the latent class model) involve (a) a draw from the PDF of the latent variables, and (b) a draw from the conditional PDF of the observed variables given the latent variables. This does not hold for LRM's, because there is no conditional PDF of observed variables given latent variables; latent variables are mapped into observed variables by means of a *function*. This mapping of latent into observed random variables is the essential new feature of LRM's, distinguishing it from classical latent variable models, and creating the possibility to formalize new substantive hypotheses like solution processes that involve multiple requirements (conjunctive) or multiple solution strategies (disjunctive).

We now consider a MCLCM that is a latent variable model in the *narrow* sense decribed above.

A compensatory model. The basic idea behind this compensatory model is that mastery of a particular mental operation may compensate for non-mastery of another mental operation. Moreover, this compensation is gradual. As such, it is different from the disjunctive LRM in which a successful application of one strategy *completely* compensates for the non-successful application of the other strategies. The compensatory model can be written as follows:

$$P(Y_i|\boldsymbol{Z};\boldsymbol{\eta}_i) = \frac{\exp\left[Y_i\left(\sum_{k=1}^{K}\eta_{ik}Z_k - \eta_{i0}\right)\right]}{1+\exp\left[\sum_{k=1}^{K}\eta_{ik}Z_k - \eta_{i0}\right]} \tag{8}$$

In this model, $K+1$ item parameters are involved. The first, η_{i0}, is a threshold parameter that specifies the value of $\sum_{k=1}^{K}\eta_{ik}Z_k$ that is needed for the probability of a correct response to be equal to 0.5. The η_{ik}-parameters are the amount of increase in probability on the log-odds scale if the person changes from non-mastery to mastery of the k-th mental operation.

This model is analogous to the item factor analysis model described by Muthèn (1978) and Bock and Aitkin (1981). The only difference is the nature of the latent variables, the Z_k's in (8). In the item factor analysis model, the latent variables are continuous, while in this model they are discrete.

In a different parametrization, the model in (8) was proposed by Hagenaars (1990, 1993) (see also Heinen, 1993) in the context of so-called causal models with discrete latent variables. Consider the model for the joint PDF of $\boldsymbol{Y}$ and $\boldsymbol{Z}$ that follows from (2), (8) and some model for $\boldsymbol{Z}$ like (1), (6) or a loglinear model. It is easy to see that this model can be written as a loglinear model for the item_1 $\times \ldots \times$item_$I\times$Classification _1 $\times \ldots \times$Classification_K-table. In this loglinear model, $-\boldsymbol{\eta}_{i0}$ is the parameter of the main effect of item i. And η_{ik} is the parameter of the interaction effect of item i and classification k. From this loglinear model for the unobservable complete table, a latent class model for the observed table of vectors $\boldsymbol{Y}$ is obtained by summing over all possible realizations of $\boldsymbol{Z}$.

Looking at (8) from the perspective of loglinear modelling leads to an interesting extension of the model. This extension involves adding three-variable interaction terms corresponding to item i and *two* classifications k and l. The additional term that appears in the exponent of (8) is then $\eta_{ikl}Z_k Z_l$. This model can be interpreted in terms of a solution process in which a correct response depends on a *joint* mastery of two mental operations. This dependence on joint mastery has some conjunctive flavor but the model is nevertheless compensatory because of the linear combination of η-parameters in the exponent.

This compensatory model will not be considered any further. Only for the MCLCM's of the LRM-type, a ML and MAP estimation algorithm will be described. The algorithm for the compensatory model, however, has essentially the same structure. ML estimation of this model is also described by Hagenaars (1990).

3 Identifiability

It can be shown that the conjunctive and the disjunctive MCLCM's presented above are not identifiable. Some identifiability restrictions are necessary to get unique, and therefore interpretable, parameter values. This non-identifiability is of the same type for the conjunctive and the disjunctive version of this model. Therefore, only the conjunctive version is considered in detail. Moreover, this non-identifiability is of the same type for all values of $K \geq 2$. Therefore, only the case $K = 2$ is considered in detail.

Two types of non-identifiability will be considered. The first one is with respect to the $P(Y_i|\boldsymbol{Z};\boldsymbol{\eta}_i)$'s and the second one is with respect to the $P(\boldsymbol{Y}:\boldsymbol{\eta},\boldsymbol{\xi})$'s, after some identifiability restrictions are imposed on the $P(Y_i|\boldsymbol{Z};\boldsymbol{\eta}_i)$'s. The $P(Y_i|\boldsymbol{Z}:\boldsymbol{\eta}_i)$'s are considered first. The $P(Y_i|\boldsymbol{Z};\boldsymbol{\eta}_i)$'s for the conjunctive LRM with $K = 2$ are shown in Figure 2. From this figure, it can be seen that there exists a multiplicative trade-off between the parameters of the two components. In particular, if η_{i11} and η_{i10} are multiplied by some constant c and η_{i21} and η_{i20} are divided by the same constant, the $P(Y_i|\boldsymbol{Z}:\boldsymbol{\eta}_i)$'s remain the same.

By fixing one of these four parameters, this multiplicative trade-off is no longer possible. We put η_{i11} equal to 1. This restriction is consistent with our interpretation of η_{i11} as the conditional probability of successful application given mastery, because it is at least as large as η_{i10}, the conditional probability given non-mastery.

		Y_i	
		0	1
$\boldsymbol{Z}$	$(0,0)$	$1-\eta_{i10}\eta_{i20}$	$\eta_{i10}\eta_{i20}$
	$(0,1)$	$1-\eta_{i10}\eta_{i21}$	$\eta_{i10}\eta_{i21}$
	$(1,0)$	$1-\eta_{i11}\eta_{i20}$	$\eta_{i11}\eta_{i20}$
	$(1,1)$	$1-\eta_{i11}\eta_{i21}$	$\eta_{i11}\eta_{i21}$

Figure 2
Conditional probabilities of the two possible realizations of Y_i according to the conjunctive LRM with $K = 2$ conditional on the four possible **Z**-patterns

It is easy to show that for $K > 2$ more than one parameter has to be fixed to make this multiplicative trade-off impossible. A restriction that is sufficient to achieve this is putting the first $(K-1)$ η_{ik1}'s equal to 1.

We now consider the second type of non-identifiability. Imposing the restrictions above is not sufficient to get an identifiable model for $\boldsymbol{Y}$. This can be shown from the formula for $P(\boldsymbol{Y};\boldsymbol{\eta},\boldsymbol{\xi})$. This formula is considered for the case of a conjunctive LRM with $K=2$ for $\boldsymbol{Y}$ given $\boldsymbol{Z}$ and the independence model for $\boldsymbol{Z}$. (The model we choose for $\boldsymbol{Z}$ is not essential for the point to be shown.)

$$
\begin{aligned}
P(\boldsymbol{Y};\boldsymbol{\eta},\boldsymbol{\xi}) &= \prod_{i=1}^{I}(\eta_{i21})^{Y_i}(1-\eta_{i21})^{1-Y_i}\xi_1\xi_2 + \\
&\quad \prod_{i=1}^{I}(\eta_{i20})^{Y_i}(1-\eta_{i20})^{1-Y_i}\xi_1(1-\xi_2) + \\
&\quad \prod_{i=1}^{I}(\eta_{i10}\eta_{i21})^{Y_i}(1-\eta_{i10}\eta_{i21})^{1-Y_i}(1-\xi_1)\xi_2 + \\
&\quad \prod_{i=1}^{I}(\eta_{i10}\eta_{i20})^{Y_i}(1-\eta_{i10}\eta_{i20})^{1-Y_i}(1-\xi_1)(1-\xi_2) \qquad (9)
\end{aligned}
$$

In this formula, the restriction that η_{i11} equals 1 is already imposed. It is possible to find another set of parameter values η^*_{i21}, η^*_{i20}, η^*_{i10}, ξ^*_1 and ξ^*_2 that results in exactly the same value as $P(\boldsymbol{Y};\boldsymbol{\eta},\boldsymbol{\xi})$ in (9). The possibility of such a transformation is best seen by first interchanging the second and the third row on the right-hand side of (9) and then performing the following transformation: $\eta^*_{i21} = \eta_{i21}$, $\eta^*_{i20} = \eta_{i10}\eta_{i21}$, $\eta^*_{i10} = \eta_{i20}/\eta_{i21}$, $\xi^*_1 = \xi_2$ and $\xi^*_2 = \xi_1$.

By imposing the restriction that η_{i21} equals 1, this transformation is reduced to a *permutation* of component item parameters ($\eta^*_{i20} = \eta_{i10}$ and $\eta^*_{i10} = \eta_{i20}$). Fortunately, from the point of view of interpretation such a permutation of parameters does not create any problems. Formally, this permutation of component item parameters is analogous to a permutation of the dimensions (factor loadings) in the factor analysis model.

For $K > 2$, the first type of non-identifiability was made impossible by putting the first $(K-1)$ η_{ik1}'s equal to 1. To remove the second type of non-identifiability, also η_{iK1} is put equal to 1. The psychological interpretation of this model involves that there is no probability of a careless error: if a person masters a particular mental operation then he or she will also apply it successfully.

For the disjunctive LRM, similar identifiability restrictions have to be imposed. For this model, all K η_{ik0}'s are put equal to 0 instead of all K η_{ik1}'s put equal to 1. The reason for this is that, in the latter case, the disjunctive LRM would predict a perfect item response pattern $\boldsymbol{Y}$ for all latent class membership patterns $\boldsymbol{z}$ different from $(0,\ldots,0)^t$.

Therefore, no distinction can be made between these latent class membership patterns on the basis of the observed item response patterns (e.g., by means of the posterior probabilities of these latent class membership patterns given the observed item response patterns). For the conjunctive LRM, an analogous problem occurs if all K η_{ik0}'s are put equal to 0 instead of all K η_{ik1}'s put equal to 1.

It is admitted that this way of dealing with the identifiability problem is not a *proof* of the model being identifiable. Evidence with respect to identifiability can also be obtained from a simulation study. In section Five, a detailed report of such a simulation study is given and the relevance of the results with respect to identifiability is pointed out.

4 Estimation

First, it is described how the EM-algorithm (Dempster, Laird and Rubin, 1977) can be used to compute the ML estimates of the MCLCM's of the LRM-type. Second, a hybrid algorithm is described that combines the EM- and a Newton-type algorithm. And third, it is described how maximum a posteriori (MAP) estimates can be obtained using the same two algorithms.

4.1 An EM-algorithm

A MCLCM is a model for the item response vectors $\boldsymbol{Y}$. For dichotomous items, there are 2^I different item response patterns. This number is denoted by S, and s ($s = 1, \ldots, S$) is used as an index for these patterns. The constant C_{si} is used to denote the response (0 or 1) on item i in pattern s. The complete vector of responses is denoted by $\boldsymbol{C}_s$ ($\boldsymbol{C}_s = (C_{s1}, \ldots, C_{sI})^t$). The number of observations of response pattern s is denoted by the random variable N_s and its realization by n_s. The vector $(N_1, \ldots, N_S)^t$ is denoted by $\boldsymbol{N}$ and its realization $(n_1, \ldots, n_S)^t$ by $\boldsymbol{n}$. The total number of observations is denoted by N_{obs}.

The following multinomial model is formulated for $\boldsymbol{N}$:

$$P(\boldsymbol{N} = \boldsymbol{n}; \boldsymbol{\eta}, \boldsymbol{\xi}) = \begin{pmatrix} N_{obs} \\ n_1 \ldots n_S \end{pmatrix} \prod_{s=1}^{S} [P(\boldsymbol{Y} = \boldsymbol{C}_s; \boldsymbol{\eta}, \boldsymbol{\xi})]^{n_s}$$

Ignoring the part of this PDF that does not depend on the parameters, the loglikelihood $l(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{n})$ can be written as follows:

$$l(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{n}) = \sum_{s=1}^{S} n_s \ln P(\boldsymbol{Y} = \boldsymbol{C}_s; \boldsymbol{\eta}, \boldsymbol{\xi})$$

The EM-algorithm can be used to maximize $l(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{n})$ because $\boldsymbol{n}$ can be considered as the *observed data* that result from a so-called *missing data generating process* in which part of the *complete data* is lost. For every person, the complete data consists of an $(I \times K)$ matrix of latent responses X_{ik}, denoted by $\boldsymbol{X}$, and a K-element vector of latent class memberships Z_k, denoted by $\boldsymbol{Z}$. This pair $(\boldsymbol{X}, \boldsymbol{Z})$ is mapped into the vector $\boldsymbol{Y}$, the observed data for this person. In the missing data generating process data are lost in two ways: (a) by applying the condensation rule to the rows of $\boldsymbol{X}$ and (b) by dropping $\boldsymbol{Z}$.

Now, some notation is introduced. For dichotomous latent responses, there are $2^{I \times K}$ different matrices $\boldsymbol{X}$. This number is denoted by R, and r $(r = 1, \ldots, R)$ is used as an index for these matrices. The constant B_{rik} is used to denote the response (0 or 1) on component k of item i in latent response matrix r. The complete matrix of latent responses is denoted by $\boldsymbol{B}_r$. Further, for dichotomous latent classifications, there are 2^K different class membership patterns. This number is denoted by T, and t $(t = 1, \ldots, T)$ is used as an index for these patterns. The constant D_{tk} is used to denote the membership (0 or 1) in classification k for pattern t. The complete vector of memberships is denoted by $\boldsymbol{D}_t$ $(\boldsymbol{D}_t = (D_{t1}, \ldots, D_{tK})^t)$. The number of observations of latent response matrix r *and* class membership vector t is denoted by the random variable M_{rt} and its realization by m_{rt}. We use $\boldsymbol{M}$ to denote the $(R \times T)$ matrix of M_{rt}'s and $\boldsymbol{m}$ to denote its realization.

The following multinomial model is formulated for $\boldsymbol{M}$:

$$P(\boldsymbol{M} = \boldsymbol{m}; \boldsymbol{\eta}, \boldsymbol{\xi}) = \begin{pmatrix} N_{obs} \\ m_{11} \ldots m_{RT} \end{pmatrix} \prod_{r=1}^{R} \prod_{t=1}^{T} [P(\boldsymbol{X} = \boldsymbol{B}_r, \boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\eta}, \boldsymbol{\xi})]^{m_{rt}}$$

Ignoring the part of this PDF that does not depend on the parameters and replacing $P(\boldsymbol{X} = \boldsymbol{B}_r, \boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\eta}, \boldsymbol{\xi})$ by the product $P(\boldsymbol{X} = \boldsymbol{B}_r | \boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\eta}) \times P(\boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\xi})$, the complete data loglikelihood $h(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{m})$ can be written as follows:

$$h(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{m}) = \sum_{t=1}^{T} m_{+t} \ln P(\boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\xi}) + \sum_{r=1}^{R} \sum_{t=1}^{T} m_{rt} \ln P(\boldsymbol{X} = \boldsymbol{B}_r | \boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\eta}), \tag{10}$$

in which m_{+t} denotes the sum over r of the m_{rt}'s for some value of t.

In the $(p+1)$-th cycle of the EM-algorithm, one maximizes the conditional expected value of the complete data loglikelihood $h(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{M})$ given the observed data $\boldsymbol{n}$ and the parameter values of the p-th cycle, denoted by $\boldsymbol{\eta}^{(p)}$ and $\boldsymbol{\xi}^{(p)}$. This function is denoted by $Q(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{\eta}^{(p)}, \boldsymbol{\xi}^{(p)})$, and it is defined as follows:

$$Q(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{\eta}^{(p)}, \boldsymbol{\xi}^{(p)}) = \mathcal{E}\left[h(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{M}) | \boldsymbol{N} = \boldsymbol{n}; \boldsymbol{\eta}^{(p)}, \boldsymbol{\xi}^{(p)}\right] \tag{11}$$

Maximizing $Q(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{\eta}^{(p)}, \boldsymbol{\xi}^{(p)})$ is simple because $h(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{m})$ is linear in the data, the m_{rt}'s (see (10)). This involves that one only has to replace the m_{rt}'s by their conditional expected values (the *E-step* of the EM-algorithm) and apply a maximization algorithm to the complete data loglikelihood (the *M-step* of the EM-algorithm). (Further, it will be shown that it is not necessary to compute expected values of *individual* M_{rt}'s but only of certain linear combinations of M_{rt}'s.)

The maximization of the complete data loglikelihood is a standard problem. First, we consider the maximization with respect to $\boldsymbol{\xi}$. Replacing $P(\boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\xi})$ in the right-hand side of (10) by the completely saturated model in (1), it can easily be shown that the ML estimates $\hat{\xi}_t$ $(t = 1, \ldots, T)$ are given by the following closed-form expression:

$$\hat{\xi}_t = \frac{m_{+t}}{N_{obs}} \quad , \tag{12}$$

the proportion of persons having class membership pattern t. And if the model for $\boldsymbol{Z}$ is the independence model in (2), it is easily shown that the ML estimates $\hat{\xi}_k$ $(k = 1, \ldots, K)$ are given by the following closed-form expression:

$$\hat{\xi}_k = \frac{\sum_{t=1}^{T} m_{+t} D_{tk}}{N_{obs}} \quad , \tag{13}$$

the proportion of persons belonging to latent class 1 of the k-th classification (having $Z_k = 1$). In a loglinear model for $\boldsymbol{Z}$, the ML estimates are the solution to a set of nonlinear equations (see Bishop, Fienberg & Holland, 1975). In these equations, the observed frequencies in some marginal tables of the complete cross-classification of the K Z_k's are put equal to their corresponding predicted frequencies. These predicted frequencies are obtained by summing the appropriate $P(\boldsymbol{Z} = \boldsymbol{D}_t; \hat{\boldsymbol{\xi}})$'s, in which $\hat{\boldsymbol{\xi}}$ is the ML estimate of $\boldsymbol{\xi}$. These equations can be solved by means of iterative proportional fitting or Newton-Raphson (see Bishop et al., 1975).

The maximization of the complete data loglikelihood with respect to $\boldsymbol{\eta}$ involves the maximization of the second term on the right-hand sided of (10). Again, this maximization is a standard problem. Because the $(I \times K)$ latent responses X_{ik} are statistically independent conditional on $\boldsymbol{Z}$, $P(\boldsymbol{X} = \boldsymbol{B}_r | \boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\eta})$ can be written as follows:

$$P(\boldsymbol{X} = \boldsymbol{B}_r | \boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\eta}) = \prod_{i=1}^{I} \prod_{k=1}^{K} P(X_{ik} = B_{rik} | Z_k = D_{tk}; \boldsymbol{\eta}_{ik})$$

Replacing $P(X_{ik} = B_{rik} | Z_k = D_{tk}; \boldsymbol{\eta}_{ik})$ by the right-hand side of (4), the second term on the right-hand side of (10) can be written as follows:

$$\begin{aligned}
&\sum_{r=1}^{R} \sum_{t=1}^{T} m_{rt} \ln P(\boldsymbol{X} = \boldsymbol{B}_r | \boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\eta}) = \\
&\quad \sum_{i=1}^{I} \sum_{k=1}^{K} \ln \eta_{ik1} \left[\sum_{r=1}^{R} \sum_{t=1}^{T} m_{rt} B_{rik} D_{tk} \right] \\
&\qquad + \ln(1 - \eta_{ik1}) \left[\sum_{t=1}^{T} m_{+t} D_{tk} - \sum_{r=1}^{R} \sum_{t=1}^{T} m_{rt} B_{rik} D_{tk} \right] \\
&\qquad + \ln \eta_{ik0} \left[\sum_{r=1}^{R} \sum_{t=1}^{T} m_{rt} B_{rik} (1 - D_{tk}) \right] \\
&\qquad + \ln(1 - \eta_{ik0}) \left[N_{obs} - \sum_{t=1}^{T} m_{+t} D_{tk} - \sum_{r=1}^{R} \sum_{t=1}^{T} m_{rt} B_{rik} (1 - D_{tk}) \right]
\end{aligned} \tag{14}$$

From (14), it can easily be shown that the ML estimates $\hat{\eta}_{ik1}$ and $\hat{\eta}_{ik0}$ are given by the following expressions:

$$\hat{\eta}_{ik1} = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T} m_{rt} B_{rik} D_{tk}}{\sum_{t=1}^{T} m_{+t} D_{tk}} \tag{15}$$

$$\hat{\eta}_{ik0} = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T} m_{rt} B_{rik}(1 - D_{tk})}{N_{obs} - \sum_{t=1}^{T} m_{+t} D_{tk}} \tag{16}$$

We now consider the E-step of the EM-algorithm. In this step, we replace the statistics that appear in $h(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{m})$ by their conditional expected values given $\boldsymbol{N} = \boldsymbol{n}$, the observed data, and $\boldsymbol{\eta}^{(p)}$ and $\boldsymbol{\xi}^{(p)}$, some preliminary values of $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$. These statistics are (a) the numerators of (12) and (13), (b) the observed frequencies in the likelihood equations for a loglinear model, and (c) the numerators and denominators of (15) and (16). The conditional expected value of the m_{+t}'s are considered first. These m_{+t}'s appear in (12), (13), (15) and (16), and can be used to compute the observed frequencies in the likelihood equations for a loglinear model.

$$\mathcal{E}\left(M_{+t} | \boldsymbol{N} = \boldsymbol{n}; \boldsymbol{\eta}^{(p)}, \boldsymbol{\xi}^{(p)}\right) = \sum_{r=1}^{R} \mathcal{E}\left(M_{rt} | \boldsymbol{N} = \boldsymbol{n}; \boldsymbol{\eta}^{(p)}, \boldsymbol{\xi}^{(p)}\right) \tag{17}$$

For every observed item response pattern $\boldsymbol{Y} = \boldsymbol{C}_s$ there is a certain probability (possibly 0 or 1) of the pattern $(\boldsymbol{X} = \boldsymbol{B}_r, \boldsymbol{Z} = \boldsymbol{D}_t)$. Therefore, the expected value in the right-hand side of (17) can be written as follows:

$$\begin{aligned}
&\mathcal{E}\left(M_{rt} | \boldsymbol{N} = \boldsymbol{n}; \boldsymbol{\eta}^{(p)}, \boldsymbol{\xi}^{(p)}\right) \\
&= \sum_{s=1}^{S} n_s P\left(\boldsymbol{X} = \boldsymbol{B}_r, \boldsymbol{Z} = \boldsymbol{D}_t | \boldsymbol{Y} = \boldsymbol{C}_s; \boldsymbol{\eta}^{(p)}, \boldsymbol{\xi}^{(p)}\right) \\
&= \sum_{s=1}^{S} n_s P\left(\boldsymbol{X} = \boldsymbol{B}_r | \boldsymbol{Y} = \boldsymbol{C}_s, \boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\eta}^{(p)}\right) P\left(\boldsymbol{Z} = \boldsymbol{D}_t | \boldsymbol{Y} = \boldsymbol{C}_s; \boldsymbol{\eta}^{(p)} \boldsymbol{\xi}^{(p)}\right)
\end{aligned} \tag{18}$$

Inserting (18) in (17) and interchanging the summations, one has to take the sum over r of $P\left(\boldsymbol{X} = \boldsymbol{B}_r | \boldsymbol{Y} = \boldsymbol{C}_s, \boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\eta}^{(p)}\right)$. This sum is equal to 1. Therefore, one gets:

$$\mathcal{E}\left(M_{rt} | \boldsymbol{N} = \boldsymbol{n}; \boldsymbol{\eta}^{(p)}, \boldsymbol{\xi}^{(p)}\right) = \sum_{s=1}^{S} n_s P\left(\boldsymbol{Z} = \boldsymbol{D}_t | \boldsymbol{Y} = \boldsymbol{C}_s; \boldsymbol{\eta}^{(p)}, \boldsymbol{\xi}^{(p)}\right) \tag{19}$$

The conditional probability in the right-hand side of (19) can be computed using Bayes' theorem:

$$P\left(\boldsymbol{Z}=\boldsymbol{D}_t|\boldsymbol{Y}=\boldsymbol{C}_s;\boldsymbol{\eta}^{(p)},\boldsymbol{\xi}^{(p)}\right)= \frac{P\left(\boldsymbol{Y}=\boldsymbol{C}_s|\boldsymbol{Z}=\boldsymbol{D}_t;\boldsymbol{\eta}^{(p)}\right)P\left(\boldsymbol{Z}=\boldsymbol{D}_t;\boldsymbol{\xi}^{(p)}\right)}{\sum_{u=1}^{T}P\left(\boldsymbol{Y}=\boldsymbol{C}_s|\boldsymbol{Z}=\boldsymbol{D}_u;\boldsymbol{\eta}^{(p)}\right)P\left(\boldsymbol{Z}=\boldsymbol{D}_u;\boldsymbol{\xi}^{(p)}\right)}$$

The $P\left(\boldsymbol{Y}=\boldsymbol{C}_s|\boldsymbol{Z}=\boldsymbol{D}_t;\boldsymbol{\eta}^{(p)}\right)$'s can be computed using (5) or (7), the conjunctive or the disjunctive LRM. And the $P\left(\boldsymbol{Z}=\boldsymbol{D}_t;\boldsymbol{\xi}^{(p)}\right)$'s can be computed using (1) or (6), one of the models for $\boldsymbol{Z}$.

The conditional expected value of the numerators of (15) and (16) are similar. Therefore, only (15) is considered. Using (18) and interchanging the summations, the following is obtained:

$$\mathcal{E}\left(\sum_{r=1}^{R}\sum_{t=1}^{T}M_{rt}B_{rik}D_{tk}|\boldsymbol{N}=\boldsymbol{n};\boldsymbol{\eta}^{(p)},\boldsymbol{\xi}^{(p)}\right)= \sum_{s=1}^{S}n_s\sum_{t=1}^{T}D_{tk}P\left(\boldsymbol{Z}=\boldsymbol{D}_t|\boldsymbol{Y}=\boldsymbol{C}_s;\boldsymbol{\eta}^{(p)},\boldsymbol{\xi}^{(p)}\right)\times \sum_{r=1}^{R}B_{rik}P\left(\boldsymbol{X}=\boldsymbol{B}_r|\boldsymbol{Y}=\boldsymbol{C}_s,\boldsymbol{Z}=\boldsymbol{D}_t;\boldsymbol{\eta}^{(p)}\right) \tag{20}$$

The sum over r on the right-hand side of (20) is the conditional probability of $X_{ik}=1$ given $Y_i=C_{si}$ and $\boldsymbol{Z}=\boldsymbol{D}_t$:

$$\sum_{r=1}^{R}B_{rik}P\left(\boldsymbol{X}=\boldsymbol{B}_r|\boldsymbol{Y}=\boldsymbol{C}_s,\boldsymbol{Z}=\boldsymbol{D}_t;\boldsymbol{\eta}^{(p)}\right)= P\left(X_{ik}=1|Y_i=C_{si},\boldsymbol{Z}=\boldsymbol{D}_t;\boldsymbol{\eta}_i^{(p)}\right) \tag{21}$$

For the conjunctive LRM, $P\left(X_{ik}=1|Y_i=C_{si},\boldsymbol{Z}=\boldsymbol{D}_t;\boldsymbol{\eta}_i^{(p)}\right)$ can be shown to be the following:

$$P\left(X_{ik}=1|Y_i=1,\boldsymbol{Z}=\boldsymbol{D}_t;\boldsymbol{\eta}_i^{(p)}\right)=1$$

$$P\left(X_{ik} = 1 | Y_i = 0, \boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\eta}_i^{(p)}\right)$$
$$= \frac{P\left(X_{ik} = 1, Y_i = 0 | \boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\eta}_i^{(p)}\right)}{P\left(Y_i = 0 | \boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\eta}_i^{(p)}\right)}$$
$$= \frac{P\left(X_{ik} = 1 | Z_k = D_{tk}; \boldsymbol{\eta}_{ik}^{(p)}\right) - P\left(Y_i = 1 | \boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\eta}_i^{(p)}\right)}{P\left(Y_i = 0 | \boldsymbol{Z} = \boldsymbol{D}_t; \boldsymbol{\eta}_i^{(p)}\right)}$$

For the disjunctive LRM, similar formula's hold.

In summary, an EM-algorithm has been presented whose E-step and M-step are both computationally feasible. In the E-step, conditional expected values are computed that involve a summation over the observed response patterns (in (18) and (20)-(21)) and a summation over the latent class membership patterns (in (20)-(21)). The number of terms in these summations is bounded above by $N_{obs} \times 2^K$. Because K is usually small, these summations are computationally feasible. And in the M-step, one has to solve a complete data maximization problem that has a closed-form solution or is solvable by means of a standard iterative algorithm.

4.2 A Hybrid Algorithm

A well-known problem of the EM-algorithm is that its convergence can be very slow (see Dempster et al, 1977). For MCLCM's this is indeed the case. Therefore, the EM-algorithm was combined with a second order algorithm (i.e., an algorithm that makes use of the matrix of second derivatives). This hybrid algorithm starts with the EM-algorithm and switches to the second order algorithm if the elements of the gradient vector are close to zero. By starting with the EM-algorithm, advantage is taken of the fact that the EM-algorithm monotonically increases the loglikelihood at all points of the parameter space. And by switching to the second order algorithm near the maximum, advantage is taken of the fact that a second order algorithm performs well at points where the surface of the loglikelihood is nearly quadratic, which is true for points near the maximum (see Gill, Murray & Wright, 1981).

The second order algorithm being used, is based on a Cholesky decomposition of the negative of the Hessian matrix (see Gill et al, 1981, pp. 108-111). This algorithm differs from the classical Newton-Raphson algorithm in that a possibly non-negative definite Hessian matrix is adjusted to make it negative definite. This allows one to construct a monotonically increasing second order algorithm (see Gill et al, 1981, pp. 99-111).

The gradient vector that is needed in this hybrid algorithm can be computed using EM-code. In particular, it is easy to show that the gradient of $l(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{n})$, the observed data loglikelihood, is equal to the conditional expected value of the gradient of $h(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{M})$, the complete data loglikelihood (see Louis, 1982). Now, since $h(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{M})$ is linear in the data, the same holds for its gradient. In particular, the gradient of $h(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{M})$ is linear in the statistics whose conditional expected values are computed in the E-step (i.e., the statistics in (12), (13), (15) and (16)). This allows one to compute the gradient of $l(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{n})$ by replacing these statistics in the gradient of $h(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{M})$ by their conditional expected values.

In the computer implementation of this hybrid algorithm, the Hessian matrix was approximated using finite differences (see Gill et al, 1981, pp. 54-56). For an exact computation, one could make use of a formula given by Louis (1982). An implementation of this formula would require new code.

The stopping criterion used in the computer implementation of this hybrid algorithm makes use of the gradient of the observed data loglikelihood. In particular, the algorithm stops if every element of this vector is less than 0.001 in absolute value.

4.3 MAP Estimation

A well-known problem in the context of models for discrete data (logistic regression, loglinear models,...) is that, depending on the data, ML estimates in the interior of the parameter space may not exist. This is also the case for MCLCM's. This fact is problematic because (a) it may result in over/underflow during computation, and (b) the asymptotic sampling properties of ML estimates (e.g., asymptotic normality) cannot be applied.

In the Bayesian framework, using a proper prior PDF, this problem does not exist. For reasons that will become clear in the following, MAP estimation will be considered here. With respect to the arbitrariness of the prior PDF, it has to be noted that, except for a constant, the likelihood function and the posterior PDF are asymptotically equivalent. Therefore, MAP and ML estimates are asymptotically equivalent.

Although ML and MAP estimates are defined in a different statistical framework, their actual computation may be very similar. In particular, the choice of a particular prior PDF in some cases is formally equivalent to adding a *prior sample* within the ML framework (see, e.g., Novick & Jackson, 1974).

4.3.1 The Beta Prior for Binomial PDF's

The independence model for $\boldsymbol{Z}$ is a product-binomial, as is the conditional PDF of the latent responses $\boldsymbol{X}$ given $\boldsymbol{Z}$. The parameters of these models, as well as the resulting models for the observed data, are all probabilities. So, not being in the interior of the parameter space involves that some parameters have boundary values of 0 or 1. For models that have probabilities as their parameters, a prior PDF that (for certain values of its parameters) is formally equivalent to a prior sample, is the *beta distribution* (see Mood, Graybill, & Boes, 1974, p.115). This PDF is defined on the domain $]0,1[$, as it should be for probabilities. The beta PDF has two parameters and if we take both of them equal to 2, this PDF can be written as follows:

$$f(W;2,2) \propto W(1-W)$$

This PDF has expected value and variance equal to 0.5 and 0.05, respectively. Letting W be any parameter of the MCLCM's being considered (ξ_k, η_{ik1} or η_{ik0}), it follows that $f(W;2,2)$ is proportional to the joint probability of a 1- and a 0-response on the latent Bernoulli random variable (Z_k or X_{ik}) whose PDF is specified by this parameter.

Making use of this prior sample interpretation of the beta prior, it is clear that MAP estimates can be computed as ML estimates using an extended sample. This extended sample involves both $\boldsymbol{N}$ and the prior sample, which will be denoted by $\boldsymbol{P}$. The array $\boldsymbol{P}$

is of order $U \times 2$, and contains one pair of observations, P_{u1} and P_{u2} for every parameter $(u = 1, \ldots, U)$. For the MCLCM's being considered, U is equal to $K + (I \times K \times 2)$ if no identifiability restrictions are imposed, and $K + (I \times K)$ if all η_{ik1}'s are fixed at 1.

The MAP estimates can be computed by means of the EM-algorithm. The complete data are $\boldsymbol{M}$ *and* $\boldsymbol{P}$. The P_{u1}'s and P_{u2}'s are considered as a special type of latent random variables that are mapped in observed random variables by means of the identity function. The function to be maximized in the M-step is $Q\left(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{\eta}^{(p)}, \boldsymbol{\xi}^{(p)}\right)$ as defined for ML estimation (see (11)), plus the loglikelihood of $\boldsymbol{P}$. This latter loglikelihood has the same structure as $Q\left(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{\eta}^{(p)}, \boldsymbol{\xi}^{(p)}\right)$. It is easy to show that the sum of $Q\left(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{\eta}^{(p)}, \boldsymbol{\xi}^{(p)}\right)$ and the loglikelihood of $\boldsymbol{P}$ is maximized by means of formula's that differ from the complete data maximizers (13), (15) and (16) in two respects: (a) the statistics are replaced by their conditional expected values, and (b) the numerators are augmented by 1 and the denominators by 2. This latter difference is due to the beta prior. Because the numerators of (13), (15) and (16) are bounded above by their denominators, it is clear that this algorithm cannot result in estimates on the boundary of the parameter space.

The second order algorithm can also be easily adapted to compute MAP estimates. In the formulas for the gradient of $l(\boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{n})$ one only has to augment the conditional expected values of certain statistics by 1 and others by 2.

4.3.2 The Dirichlet Prior for Multinomial PDF's

The saturated model for $\boldsymbol{Z}$ (see (1)) is multinomial. The parameters of this model are a vector of probabilities, restricted to sum to 1. A prior PDF that has the same structure as the multinomial PDF, is the *Dirichlet* (see Bishop et al, 1975, p. 405). With appropriate parameter values, the Dirichlet is formally equivalent to a prior sample of a number of subjects distributed over the T latent class patterns. Maximizing the posterior PDF for a MCLCM with a saturated model for $\boldsymbol{Z}$ proceeds in essentially the same way as in the case with an independence model together with a beta prior. The only difference is that the statistics that have to be augmented by values coming from the prior are the m_{+t}'s instead of the $\sum_{t=1}^{T} m_{+t} D_{tk}$'s

4.3.3 The Logistic Prior for Loglinear and Logistic Models

In the previous, we mentioned the possibility of formulating a loglinear model for $\boldsymbol{Z}$, and in (8), we presented a logistic model for $\boldsymbol{Y}$ given $\boldsymbol{Z}$. For these models, not being in the interior of the parameter space involves that some parameters have boundary values of $-\infty$ or $+\infty$. A prior PDF that has the same structure as a loglinear or a logistic model, is the *logistic PDF* (see Mood, Graybill, & Boes, 1974, p. 118). For suitable parameter values, this PDF is equivalent to a prior sample of two observations (a 1-response and a 0-response) on a hypothetical data point characterized by a probability that has a loglinear/logistic structure.

Maximizing the posterior PDF is possible by means of a simple modification of the algorithm for the maximization of $Q(\boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\eta}^{(p)}, \boldsymbol{\xi}^{(p)})$. Also, the gradient of the observed data log joint PDF can be computed by means of a simple modification of the formulas for the gradient of the observed data loglikelihood.

4.4 A Fully Bayesian Approach

A fully Bayesian approach would involve that the complete posterior PDF of the parameters given the data be computed, not only its mode. This posterior PDF is not analytically available. However, using standard asymptotic theory, it can be approximated by a multivariate normal PDF with expectation equal to the mode, and covariance matrix equal to the inverse of the Hessian matrix at the mode. For small samples, the approximation can be improved by making use of the *sampling-importance-resampling algorithm* (Rubin, 1987). This algorithm allows one to obtain draws from the posterior PDF. These draws can then be used (a) to compute the moments of this PDF, and (b) to construct 95-percent posterior probability intervals.

5 Simulation Study

In a simulation study, two properties of the MAP estimates were examined: (a) the uniqueness of the solution of the *posterior PDF equations* (i.e., the gradient of the log posterior PDF equated to zero), and (b) the goodness-of-recovery (GOR) of the true parameter values. As an extension to the GOR-study, also the existence of asymptotic standard errors was examined.

5.1 The Simulation and the Estimation Design

Data were generated according to the conjunctive MCLCM with the independence model (see (2)) as a model for $\boldsymbol{Z}$. The simulation design involves three factors: (a) the number of items (I) with two levels (I= 20 or 40), (b) the number of components (K) with two levels (K= 2 or 4) and (c) the number of persons (N) with four levels (N= 100, 250, 1000 or 2500). Thus, the simulation design has $2 \times 2 \times 4 = 16$ cells. For every cell in this design, 100 random data sets were generated, using parameter values to be described in the following.

The independence model for $\boldsymbol{Z}$ is characterized by K parameters, the marginal probabilities $P(Z_k = 1)$. For $K = 2$, these marginal probabilities are 0.6 and 0.4, and for $K = 4$, they are 0.6, 0.4, 0.7 and 0.5. Item responses were generated with all η_{ik1}-parameters ($P(X_{ik} = 1|Z_k = 1)$'s) equal to 1. This corresponds to the identifiability restriction that was imposed when doing the estimation. The values for the η_{ik0}-parameters ($P(X_{ik} = 1|Z_k = 0)$'s) are shown in Figure 3. For $K = 2$, only the first two columns were used. For items 21 to 40 in the cells with $I = 40$, the same η_{ik0}-parameters were used as for items 1 to 20.

Item	Component 1	Component 2	Component 3	Component 4
1	.9	.9	.9	.9
2	.9	.9	.9	.1
3	.9	.1	.9	.9
4	.9	.9	.1	.1
5	.1	.9	.9	.1
6	.9	.1	.1	.9
7	.1	.1	.9	.9
8	.1	.1	.1	.9
9	.1	.9	.1	.1
10	.1	.1	.1	.1

Item	Component 1	Component 2	Component 3	Component 4
11	.9	.9	.9	.9
12	.9	.9	.9	.5
13	.9	.5	.9	.9
14	.9	.9	.5	.5
15	.5	.9	.9	.5
16	.9	.5	.5	.9
17	.5	.5	.9	.9
18	.5	.5	.5	.9
19	.5	.9	.5	.5
20	.5	.5	.5	.5

Figure 3
Values of the η_{iko}-parameters used in the simulation study

By *estimation design*, we mean which analyses have been performed on which data sets. In this study, this was very simple: every data set was analyzed twice: once using the conjunctive MCLCM with the independence model for $\boldsymbol{Z}$, and once using the conjunctive MCLCM with the saturated model for $\boldsymbol{Z}$.

5.2 Uniqueness

To examine uniqueness, 40 data sets were used; five from each of the eight cells in the design with number of items equal to 20. For each of these data sets, the parameters of the conjunctive MCLCM were estimated 10 times using the independence model for $\boldsymbol{Z}$, and 10 times using the saturated model. Because every analysis started from different random starting values, these analyses give information that is relevant for the uniqueness problem.

For six out of eight cells in the design, the algorithm always converged to the same solution point. This was the case for all four 2-component data sets/analyses and the two largest ($N = 1000, 2500$) 4-component data sets/analyses, both with the independence model and the saturated model for $\boldsymbol{Z}$. For the two smallest ($N = 100, 250$) 4-component data sets/analyses, *multiple* solution points were found for most of the data sets (8 out of 10 data sets for the analyses with the independence model, and 7 out of 10 for the analyses with the saturated model). In these analyses, the solution with the highest posterior PDF was the most frequent: in more than 60 percent of the analyses the algorithm converged to the best solution. Therefore, performing multiple analyses using different random starting values may still be a successful optimization method in these cases.

For all data sets, the solution points, whether single or multiple, were allways local maxima and never saddle points or points in a linear subspace. This follows from the fact that the Hessian matrix at the solution points was allways negative definite. This was also the case for all analyses that were performed in the goodness-of-recovery study.

5.3 Goodness-of-Recovery and the Existence of Asymptotic Standard Errors

For every parameter, two GOR-statistics were computed: (a) the difference between the true parameter value and the average estimated parameter value (denoted by BIAS), and (b) the root mean square deviation between the true and the estimated parameter values (denoted by RMSD). Both averages were computed over the 100 replications in every cell. Formally, they are defined as follows:

$$\text{BIAS} = \left(\frac{\sum_{i=1}^{100} \hat{\theta}^{(i)}}{100} \right) - \theta$$

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{100}(\hat{\theta}^{(i)} - \theta)^2}{100}}$$

In these equations, i is an index for the replications, θ is some parameter and $\hat{\theta}^{(i)}$ is its estimate in replication i. The RMSD is a function of both the bias and the standard error (SE) of the estimates. The SE is estimated by the Monte Carlo standard error (MCSE):

$$\mathrm{MCSE} = \sqrt{\frac{\sum_{i=1}^{100}(\hat{\theta}^{(i)} - \bar{\theta})^2}{100}},$$

in which $\bar{\theta}$ is the average estimated parameter value. It is easy to show that

$$\mathrm{RMSD}^2 = \mathrm{MCSE}^2 + \mathrm{BIAS}^2$$

In interpreting the results, it is useful to know that, if the bias is negligable and the sampling distribution is normal, a 95-percent confidence interval has width equal to ($2 \times 1.96 \times$ RMSD). If the 95-percent confidence intervals for the latent response probabilities (the η_{ik0}'s) have an average width of less than 20 percent of the range of a probability (i.e., $(0,1)$), we say that the goodness-of-recovery is sufficient for substantive interpretation of the components.

Besides the MAP estimates, in every analysis, also the Hessian matrix of the log posterior PDF was computed. It is known that, under certain regularity conditions, the inverse of this matrix converges to the sampling covariance matrix of the MAP estimates, or, from a Bayesian perspective, the posterior covariance matrix of the parameters (see, e.g., Gelman, Carlin, Stern, & Rubin, 1995, pp. 94-111). Here, we take the frequentist perspective, considering the MAP estimates as point estimates of unknown constants, and ask the question how well this inverse Hessian matrix approximates the true sampling covariance matrix for finite, but increasing, sample sizes. For MCLCM's, as opposed to other models like the loglinear model, this question is especially important, since we have no proof of the identifiability of the MCLCM's, and identifiability is one of the regularity conditions for the asymptotic result to hold.

For a given parameter, inverting the Hessian matrix gives us one asymptotic standard error (ASE) per replication in a cell of the simulation design. These 100 ASE were aggregated by computing the mean ASE (MASE). A possible bias in the asymptotic SE's is found by computing the difference between these MASE-statistics and the MCSE's.

When computing GOR-statistics, one has to be sure that one has actually computed the estimate whose performance one wants to evaluate, in our case the MAP estimate. As is clear from the uniqueness study, for the small ($N = 100, 250$) 4-component data sets/analyses the solution of the maximization algorithm is not always the MAP estimate. Therefore, the values of the GOR-statistics in these cells of the simulation design will not be considered.

A further complication is that MCLCM's involve a trivial non-identifiability caused by the fact that the components may be permuted without changing the probabilities of the observed item response vectors. For instance, with the independence model for $\boldsymbol{Z}$, the marginal latent class probability and the I latent response probabilities of the first component (i.e., ξ_1, η_{110}, η_{210}, $\ldots$, η_{I10}) may be replaced by the corresponding probabilities for the second component, and vice versa, without changing the $P(\boldsymbol{Y}; \boldsymbol{\xi}, \boldsymbol{\eta})$'s (a similar, but somewhat more complicated, permutation is possible with the saturated model for $\boldsymbol{Z}$). To get the same order of components for each of the replications, the *true* parameter values were used as reference values. In particular, for every replication, a deviance measure was computed for every (true component, estimated component)-pair. This deviance measure is simply the sum of the absolute differences between the corresponding true and estimated latent response probabilities (the $\hat{\eta}_{ik0}$'s). Formally, for the k-th true and the l-th estimated component, this deviance measure is the following:

$$\sum_{i=1}^{I} |\eta_{ik0} - \hat{\eta}_{il0}|$$

Every estimated component was given the order of the true component from which it deviated least. It should be clear that this way of determining the order of the components only works if the goodness-of-recovery is not too bad. Otherwise, it might occur for example that one true component is the best choice for more than one estimated component. Fortunately, this turned out to be the case for none of the data sets considered. This fact of having every true component represented best in only one estimated component is already an indication of at least a reasonable goodness-of-recovery.

Only the results for the [20 items, independence model] data sets/analyses will be considered in detail. At the end, the effects of number of items (20/40) and type of model for Z (independence/saturated) will be discussed briefly. In Table 1, the results are shown for the [20 items, 2 components, 100 persons, independence model] data sets/analyses. In this table, the following is shown: the average parameter estimates, the true parameter values, and the BIAS-, RMSD-, MCSE-, and MASE-statistics. These values are given for the marginal latent class probabilities (ξ_1 and ξ_2) and the latent response probabilities (the η_{ik0}'s) of items 1, 3, 5 and 7. In this table, also the averages over *all* η_{ik0}'s (so, not only items 1, 3, 5, and 7) of these six statistics are shown. (For the BIAS-statistics, the average of their *absolute* values was computed, whereas for the others the ordinary average was computed.) First, taking into account the small number of persons (100) and the relatively large number of parameters (42), goodness-of-recovery is satisfactory. However, it is insufficient for a substantive interpretation of the components. Second, the MASE's are too large (by an average percentage of 10.96).

Table 1
GOR-statistics for the [20-items, 2-component, 100 persons, independence model] datasets/analyses (see text for explanation)

Parameter	*Average*	*True*	*BIAS*	*RMSD*	*MCSE*	*MASE*
ξ_1	0.6072	0.6	0.0072	0.0466	0.0461	0.0522
ξ_2	0.4072	0.4	0.0072	0.0555	0.0550	0.0518
η_{110}	0.8777	0.9	-0.0223	0.0541	0.0493	0.0621
η_{120}	0.8834	0.9	-0.0166	0.0458	0.0427	0.0471
η_{310}	0.8141	0.9	-0.0859	0.1180	0.0809	0.1030
η_{320}	0.1183	0.1	0.0183	0.0422	0.0380	0.0465
η_{510}	0.1199	0.1	0.0199	0.0615	0.0582	0.0594
η_{520}	0.8693	0.9	-0.0307	0.0589	0.0503	0.0608
η_{710}	0.1425	0.1	0.0425	0.0882	0.0773	0.0759
η_{720}	0.1171	0.1	0.0171	0.0466	0.0433	0.0498
Average over all item parameters	0.5909	0.6	0.0235	0.0680	0.0615	0.0675

In Table 2, the results are shown for the [20 items, 2 components, 1000 persons, independence model] data sets/analyses. First, goodness-of-recovery now is sufficient for substantive interpretation; the 95-percent confidence intervals for the latent response probabilities have an average width of about 0.083. Second, as it should be, all GOR-statistics are much smaller for the 1000- than for the 100-persons data sets/analyses. And third, also the *differences* between the MASE's and the MCSE's are much smaller than for the 100-persons data sets/analyses. The MASE's are too large by an average percentage of 1.39 only.

Table 2

GOR-statistics for the [20-item, 2-component, 1000 persons, independence model] data sets/analyses (see text for explanation)

Parameter	*Average*	*True*	*BIAS*	*RMSD*	*MCSE*	*MASE*
ξ_1	0.5997	0.6	-0.0003	0.0169	0.0169	0.0167
ξ_2	0.4026	0.4	0.0026	0.0165	0.0163	0.0163
η_{110}	0.8996	0.9	-0.0004	0.0187	0.0187	0.0195
η_{120}	0.8981	0.9	-0.0019	0.0154	0.0153	0.0149
η_{310}	0.8946	0.9	-0.0054	0.0274	0.0268	0.0296
η_{320}	0.1005	0.1	0.0005	0.0144	0.0144	0.0137
η_{510}	0.0996	0.1	-0.0004	0.0171	0.0171	0.0180
η_{520}	0.8970	0.9	-0.0030	0.0178	0.0175	0.0189
η_{710}	0.1021	0.1	0.0021	0.0226	0.0225	0.0227
η_{720}	0.1005	0.1	0.0005	0.0152	0.0152	0.0156
Average over all item parameters	0.5987	0.6	0.0028	0.0212	0.0209	0.0211

For the [20 items, 2 components, 2500 persons, independence model] data sets/analyses the values of the GOR-statistics are again smaller than for the 1000-persons data sets/analyses. In particular, the average absolute BIAS- and RMSD-values of the item parameters are 0.0012 and 0.0136, respectively. The average difference between the MCSE's and the MASE's is further reduced to an average percentage of 0.13.

We now consider the results for the [20 items, 4 components, independence model] data sets/analyses. Because the estimation algorithm does not result in unique solutions for the small ($N = 100, 250$) 4-component data sets/analyses, only the results for the larger ($N = 1000, 2500$) 4-component data sets/analyses are considered. In Table 3, the results are shown for the [20 items, 4 components, 1000 persons, independence model] data sets/analyses (for items 1, 3, 5, 7 and 10). First, goodness-of-recovery is sufficient for substantive interpretation of the components: the 95-confidence intervals for the latent response probabilities have an average width of about 0.143. Second, comparing the results in Table 3 with those for the corresponding 2-component data sets/analyses in Table 2, we see that estimating more parameters is at the expense of a goodness-of-recovery: the average RMSD is 33 percent larger for the latent class membership probabilities and 73 percent for the latent response probabilities. Third, there is an interesting relation between the four marginal latent class probabilities and the goodness-of-recovery of the corresponding latent response probabilities: the *larger* the marginal latent class probability, the *better* the goodness-of-recovery of the corresponding latent response probabilities. In particular, the four marginal latent class probabilities are 0.6, 0.4, 0.7, and 0.5, and the average (over the items) RMSD's of the corresponding latent response probabilities are 0.0388, 0.0266, 0.0498, and 0.0314, respectively. (The same pattern was found in all 4-component data sets/analyses.) This pattern is understandable because only the persons with $Z_k = 0$ give information about the η_{ik0}'s. And fourth, as is the case for the 2-components data sets/analyses, the MASE's are on the average (over the parameters) larger than the MCSE's (in particular, by an average percentage of 5.38).

Table 3

GOR-statistics for the [20-items, 4-component, 1000 persons, independence model] data sets/analyses (see text for explonation)

Parameter	*Average*	*True*	*BIAS*	*RMSD*	*MCSE*	*MASE*
ξ_1	0.6000	0.6	0.0000	0.0220	0.0220	0.0253
ξ_2	0.4116	0.4	0.0116	0.0224	0.0191	0.0203
ξ_3	0.6921	0.7	-0.0079	0.0224	0.0210	0.0260
ξ_4	0.5094	0.5	0.0094	0.0221	0.0200	0.0216
η_{110}	0.8908	0.9	-0.0092	0.0291	0.0276	0.0290
η_{120}	0.8968	0.9	-0.0032	0.0222	0.0220	0.0212
η_{130}	0.8952	0.9	-0.0048	0.0324	0.0321	0.0337
η_{140}	0.9044	0.9	-0.0044	0.0234	0.0230	0.0235
η_{310}	0.8834	0.9	-0.0166	0.0450	0.0418	0.0444
η_{320}	0.0979	0.1	-0.0021	0.0168	0.0167	0.0188
η_{330}	0.8589	0.9	-0.0411	0.0679	0.0541	0.0573
η_{340}	0.8934	0.9	-0.0066	0.0311	0.0304	0.0314
η_{510}	0.1009	0.1	0.0009	0.0327	0.0327	0.0316
η_{520}	0.8856	0.9	-0.0144	0.0344	0.0312	0.0321
η_{530}	0.8644	0.9	-0.0356	0.0654	0.0539	0.0615
η_{540}	0.0993	0.1	-0.0007	0.0202	0.0202	0.0225
η_{710}	0.1073	0.1	0.0073	0.0369	0.0362	0.0335
η_{720}	0.1009	0.1	0.0009	0.0194	0.0193	0.0197
η_{730}	0.8520	0.9	-0.0480	0.0791	0.0629	0.0645
η_{740}	0.8879	0.9	0.0121	0.0388	0.0369	0.0386
η_{1010}	0.1053	0.1	0.0053	0.0360	0.0356	0.0373
η_{1020}	0.1029	0.1	0.0029	0.0270	0.0268	0.0252
η_{1030}	0.1135	0.1	0.0135	0.0508	0.0490	0.0467
η_{1040}	0.1056	0.1	0.0056	0.0315	0.0310	0.0306
Average over all item parameters	0.5945	0.6	0.0089	0.0366	0.0349	0.0365

For the [20 items, 4 components, 2500 persons, independence model] data sets/analyses the values of all GOR-statistics are smaller than for the 1000-persons data sets/analyses, similar to the difference between the results for the corresponding 2-components data sets/analyses.

The effect of the number of items on the goodness-of-recovery is rather small. Doubling the number of items (i.c., 40 instead of 20) resulted in a global decrease of the RMSD's of the marginal latent class (pattern) probabilities by an average percentage of 14.51, and of the latent response probabilities by an average percentage of 9.42.

The effect of the model for $\boldsymbol{Z}$ (i.c., independence versus saturated) on the goodness-of-recovery of the latent response probabilities is negligable. Under the saturated model, the RMSD's of the latent response probabilities are smaller than under the independence model by an average percentage of 0.67 only. For the parameters of the model for $\boldsymbol{Z}$, we start from the observation that the independence model has two or four ($K = 2, 4$) free parameters, and the corresponding saturated model three, respectively, fifteen (that is, $2^K - 1$). For the 2-component data sets/analyses, the RMSD's of the parameters of the saturated model are *smaller* than those of the independence model by an average percentage of 15.53. For the 4-component data sets/analyses this average percentage is 74.76. This result is somewhat puzzling, since the model with the larger number of parameters has a better goodness-of-recovery.

From this goodness-of-recovery study we can draw four conclusions. First, goodness-of-recovery was more than satisfactory, even with as few as 100 persons. However, for substantive interpretation of the components, more persons are required. Second, the global decrease of the RMSD's with increasing sample size strongly suggests consistent estimation. Third, there is good evidence that the diagonal elements of the inverse of the Hessian matrix of the log posterior PDF are asymptotic sampling variances; although the elements are systematically larger than the true sampling variances. this bias disappears with increasing sample size. Fourth, the positive results with respect to goodness-of-recovery, and the fact that the usual asymptotic standard errors appear to be valid, are indirect evidence for the identifiability of the conjunctive MCLCM (at least for the parameter values used in the simulation study). However, it is clear that only an analytic identifiability proof, probably involving conditions on the model structure and the parameter values, is fully satisfactory. Proving identifiability of MCLCM's would be a major step forward, since it would give us a class of discrete latent variable models that does not suffer from rotational invariance (or something similar), as does the factor analysis model.

6 Application

The conjunctive MCLCM (with identification restrictions as described in section 3) was applied to data collected by Tatsuoka (1984). The items are fraction addition problems. The item set consisted of both simple (e.g., $3/4 + 1/2 =$, $1/5 + 1/4 =$) and mixed (e.g., $3\ 5/7 + 4\ 6/7 =$, $1\ 1/3 + 2\ 4/6 =$) fraction addition problems. The test consists of 38 items. The first half of the test is parallel to the second half. This means that for every item in the first half, there is another item of the same type (single/mixed, same/different denominator, small/large numbers, ...) in the second half. Item 26 was excluded from the analysis because the information about this item in the documentation (i.e., Tatsuoka's (1984) report) was not consistent with item 26 in the list that was also part of this documentation. There were 595 subjects. In this group, there were children from grade 7, 8 and 9.

The conjunctive MCLCM, instead of the disjunctive or some other type, was chosen because the solution of this type of items can be described well in terms of multiple abilities that are *all* necessary for a correct response (see the interpretation below). Because we had no hypotheses about the structure in the latent class membership probabilities, the saturated model was chosen.

We obtained MAP-estimates for the conjunctive MCLCM with from one to five components. The analyses were performed using multiple random starting-values. Except for the 4- and the 5-component model, these analyses always resulted in the same parameter values. Both for the 4- and the 5-component model, two local maxima were found. Only the solutions with the largest value of the posterior density are considered in the following. In Table 4, we give (a) the number of parameters, and (b) the value of the loglikelihood at the MAP-estimates, for each of these five solutions. Note that the function being maximized is not the loglikelihood, but the loglikelihood *plus* the log prior density. One might consider using the likelihood ratio statistic to determine the number of components. Unfortunately, likelihood ratio statistics for mixture models involving different numbers of latent classes do not have the usual asymptotic chi-square distribution (with degrees of freedom equal to the difference between the number of parameters of the two models; see McLachlan & Basford, 1988). Therefore, the loglikelihood was used as a descriptive statistic only. A scree-plot of the loglikelihood values shows that the decrease in loglikelihood is almost linear from the 3- to the 5-component solution.

Table 4

Number of parameters and loglikelihood values for the conjunctive MCLCM with from one to five components applied to Tatsuoka'sfraction addition data

Number of Components	*Number of Parameters*	*Loglikelihood at the MAP-estimates*
1	39	-14810.70737
2	79	-9321.21454
3	121	-7990.82645
4	167	-7616.34449
5	221	-7452.08613

Besides the scree-plot, a second reason for restricting our attention to the 3-component solution, is that the interpretation of the 4-component solution is along the same lines as the simpler interpretation of the 3-component solution. This is evident from the pattern of correlations between the latent response probabilities of the different components and two item characteristics (see further).

The 3-component solution is given in Tables 5 and 6. The estimates and the associated standard errors of the ξ_t's ($t = 1, \ldots, 8$) are given in Table 5 (for identifiability, ξ_1 is put equal to 1 minus the sum of the other ξ_t-parameters). The estimates and the associated standard errors of the latent response probabilities (the η_{ik0}'s) of the first 19 items are given in Table 6. The averages in the bottom of this table are over all 37 items. In this table, we also show two item characteristics that are used for the validation of the interpretation of this solution (see further).

For the interpretation, it is important to know that (a) η_{ik0} is the probability of a correct application of the k-th mental operation, given that one does not master it ($Z_k = 0$), and (b) subjects with $Z_k = 1$ allways apply the k-th mental operation correctly. The latter restriction was introduced for identification of the model (see Section 3).

Table 5

MAP-estimates of the latent class pattern probabilities under the 3-component conjunctive MCLCM (with a saturated model for **Z**) applied to Tatsuoka's fraction addition data

Pattern	*Estimate [St. Err.]*
000	0.2204 Id. Res.
100	0.0838 [0.0162]
010	0.0075 [0.0062]
110	0.5232 [0.0204]
001	0.0237 [0.0220]
101	0.1048 [0.0168]
011	0.0234 [0.0082]
111	0.0128 [0.0046]

Table 6

MAP-estimates of the lataent response probabilities under the 3-component conjunctive MCLCM applied to Tatsuoka's fraction addition data

Item	*LRP Comp. 1*	*LRP Comp. 2*	*LRP Comp. 3*	*Com./Diff. Denom.*	*STM Load*
$2\,\frac{8}{6} + 3\,\frac{10}{6}$	0.3157 [0.0614]	0.6096 [0.0556]	0.7729 [0.0240]	1	2
$\frac{2}{5} + \frac{12}{8}$	0.4117 [0.1275]	0.0778 [0.0255]	0.6662 [0.0266]	0	7
$\frac{8}{5} + \frac{6}{5}$	0.1698 [0.0429]	0.6490 [0.0521]	0.8788 [0.0186]	1	2
$2\,\frac{1}{2} + 4\,\frac{2}{4}$	0.3662 [0.1135]	0.0960 [0.0270]	0.8342 [0.0210]	0	3
$\frac{1}{2} + 1\,\frac{10}{7}$	0.2495 [0.1042]	0.0397 [0.0189]	0.6289 [0.0272]	0	5
$3\,\frac{5}{7} + 4\,\frac{6}{7}$	0.2046 [0.0443]	0.7508 [0.0536]	0.8062 [0.0228]	1	3
$\frac{3}{5} + \frac{7}{5}$	0.1220 [0.0357]	0.6273 [0.0531]	0.8950 [0.0177]	1	1
$\frac{1}{3} + \frac{1}{2}$	0.5917 [0.1253]	0.0234 [0.0118]	0.9106 [0.0162]	0	3
$1\,\frac{4}{7} + 1\,\frac{12}{7}$	0.2602 [0.0596]	0.5537 [0.0548]	0.7792 [0.0237]	1	3
$\frac{3}{5} + \frac{1}{5}$	0.1591 [0.0343]	0.8198 [0.0400]	0.9766 [0.0087]	1	1
$\frac{3}{4} + \frac{1}{2}$	0.4097 [0.1222]	0.0665 [0.0221]	0.8922 [0.0174]	0	3
$2\,\frac{5}{9} + 1\,\frac{1}{9}$	0.2749 [0.0526]	0.6502 [0.0518]	0.8959 [0.0174]	1	3
$3\,\frac{1}{6} + 2\,\frac{3}{4}$	0.3046 [0.1169]	0.0204 [0.0131]	0.8190 [0.0217]	0	5
$\frac{15}{35} + \frac{10}{35}$	0.1714 [0.0410]	0.7505 [0.0460]	0.9084 [0.0164]	1	3
$\frac{1}{2} + \frac{3}{8}$	0.3499 [0.1204]	0.0191 [0.0124]	0.9260 [0.0148]	0	2
$1\,\frac{2}{5} + \frac{3}{5}$	0.1041 [0.0287]	0.7781 [0.0458]	0.9403 [0.0136]	1	1
$\frac{1}{4} + \frac{3}{4}$	0.1420 [0.0305]	0.8961 [0.0351]	0.9579 [0.0113]	1	1
$\frac{4}{15} + \frac{1}{10}$	0.5364 [0.1319]	0.0116 [0.0094]	0.8001 [0.0225]	0	4
$\frac{4}{5} + \frac{3}{5}$	0.0968 [0.0271]	0.8722 [0.0417]	0.9113 [0.0165]	1	2
Average over all item parameters	0.2023 [0.0638]	0.4329 [0.0340]	0.8418 [0.0193]		

The first component both has the largest probability of being mastered (i.e., 0.72) and a uniformly small probability of compensation by other mental resources (around 0.20). Therefore, the corresponding latent ability most likely involves a mental operation that is involved in every item. This is true for the mental operation of adding simple fractions with a common denominator (i.e., $a/c + b/c = (a+b)/c$) since this operation has to be performed in every item.

The second component involves knowing how fractions with different denominators have to be converted in equivalent fractions involving a common denominator. This is convincingly demonstrated by the 0.97 correlation between the latent response probabilities and the binary variable indicating whether the item has common (value 1) or different (value 0) denominators. This means that items with common denominators do not require this ability, whereas items with different denominators do require it.

The third component involves short-term memory (STM) capacity. To show this, we first describe which elements have to be stored in STM. In particular, solving fraction addition problems involves that the following numbers are computed and stored temporarily in STM:

- (If there are two unit-parts involved.) The sum of the unit-parts.
- (If the denominators of the fractions are common.) The sum of the numerators.
- (If the denominators of the fractions are different.)

    1. The new common denominator, which is determined as the smallest number that is a multiple of both old denominators (e.g., 12 for the old denominators 6 and 4).
    2. The two new numerators, obtained by multiplying the old numerator by the same factor that was used to convert the denominator.
    3. The sum of the two new numerators, which replaces the two numerators in STM.

- (If the sum of the fractions can be simplified.) The greatest common divisor of numerator and denominator.

For the validation of our hypothesis that the third component involves STM capacity, we compute the STM load of each fraction addition problem. This STM load is computed as the largest number of digits that is at the same time in STM. Computing the number of *digits* instead of the number of *numbers* in STM, reflects the assumption that it is more difficult to store a two-digit number than a single-digit number. Our hypothesis is corroborated by the negative correlation of -0.83 between STM load and the latent response probabilities of the third component. This means that items with a high STM load can only be solved by subjects having $Z_3 = 1$. Further, as a kind of discriminant validity, these latent response probabilities are only weakly correlated with the binary variable indicating common versus different denominators (i.c., 0.40).

Since this interpretation is post-hoc, a caveat is in order. A confirmatory study is needed to give this theory a more solid basis. Such a study might involve a test for STM capacity that is correlated with the a posteriori probabilities of belonging to the mastery-class of the STM component (as determined by the correlations between the latent response probabilities and STM load).

7 Related Models

MCLCM's are a generalization of ordinary latent class models because they involve *multiple* latent classifications instead of only *one*. This generalization is also presented by Hagenaars (1990, 1993) using a loglinear model for the PDF of the item responses conditional on the latent class memberships (see 2.2.2., the compensatory model). From the perspective of latent class modelling, MCLCM's with *polytomous* instead of dichotomous latent classifications are a straightforward generalization. Obviously, polytomous latent classifications cannot be interpreted as mastery/non-mastery of mental operations. However, they may be useful in modelling responses to items in the personality and clinical domain as, for example, items in a psychiatric symptom checklist. For this type of responses it makes more sense a priori to think in terms of polytomous latent classifications as, for example, [conforming, independent, antagonistic] as a relevant classification for attitude towards rules.

The MCLCM's of the LRM-type are related to a family of decomposition models for binary two-way data. The first model in this family is the so-called *HICLAS model* of De Boeck and Rosenberg (1988) that involves a Boolean matrix decomposition. Boolean matrix decomposition is the deterministic limiting case of the disjunctive MCLCM. To show this, consider the disjunctive MCLCM with all η_{ik0}-parameters equal to 0. The remaining item parameters (the η_{ik1}'s) can then be written in a K-dimensional vector $\boldsymbol{\eta}_i = (\eta_{i11}, \ldots, \eta_{iK1})^t$. Then, the probability of a correct response conditional on the latent class memberships can be written as follows:

$$P(Y_i = 1|\boldsymbol{Z}; \boldsymbol{\eta}_i) = 1 - \prod_{k=1}^{K}(1 - \eta_{ik1} Z_k) \tag{22}$$

The deterministic limiting case of this probability is obtained by letting the elements of $\boldsymbol{\eta}_i$ go to either 0 or 1. In the limit, the right-hand side of (27) is the Boolean vector product of the binary vectors $\boldsymbol{\eta}_i$ and $\boldsymbol{Z}$. This Boolean vector product has the value 1 if and only if there is at least one k for which $\eta_{ik1} = Z_k = 1$.

A conjunctive version of the HICLAS model was proposed by Van Mechelen, De Boeck and Rosenberg (1995). This version of the model involves a kind of conjunctive Boolean matrix decomposition. This conjunctive Boolean matrix decomposition is the limiting case of the conjunctive MCLCM, as can be shown in a similar way as for the original Boolean matrix decomposition and the disjunctive MCLCM.

It should be noted that the HICLAS model is not a latent class model. The $\boldsymbol{Z}$'s are not considered as RV's but as constants that are treated as unknown parameters (one for each of the persons in the sample). Thus, in an application of the HICLAS model, $N + I$ binary vectors of constants (N $\boldsymbol{Z}$'s and I $\boldsymbol{\eta}_i$'s) have to be estimated from the data.

Probability matrix decomposition (PMD) models were proposed by Maris et al. (1996) as probabilistic versions of the Boolean matrix decompositions involved in the two versions of the HICLAS model. PMD models are introduced here using a matrix of patients by psychiatric symptoms. For every cell in this matrix, a number of psychiatrists indicated whether the corresponding symptom applied to the corresponding patient. Thus, the data matrix is a matrix of frequencies instead of 0's and 1's. Both the patients and the symptoms are characterized by a K-dimensional vector of probabilities. For the symptoms, this is similar to a MCLCM. There also, the items are characterized by a K-dimensional vector of probabilities (after imposing identifiability restrictions on the η_{ik0}'s or the η_{ik1}'s). The difference between PMD-models and MCLCM's is the representation of the persons: in PMD models, this representation is a vector of probabilities, and in MCLCM's, it is a realization of a vector-valued binary RV.

To give a formal characterization of PMD models, the vector of probabilities for patient p is denoted by $\boldsymbol{\zeta}_p = (\zeta_{p1}, \ldots, \zeta_{pK})^t$. And the binary RV indicating whether symptom i was judged to be applicable to patient p is denoted by Y_{pi}. Then, analogous to (27), the probability of $Y_{pi} = 1$ under the disjunctive PMD model can be written as follows:

$$P(Y_i = 1|\boldsymbol{Z}; \boldsymbol{\zeta}_p, \boldsymbol{\eta}_i) = 1 - \prod_{k=1}^{K}(1 - \eta_{ik1}\zeta_{pk}) \tag{23}$$

A similar formula holds for the probability of $Y_{pi} = 1$ under the conjunctive PMD model.

The disjunctive and the conjunctive MCLCM are half-way between the HICLAS and the PMD models. Ignoring the fact that $\boldsymbol{Z}$ is a RV in the MCLCM's, we go from the MCLCM's to the HICLAS model by replacing the vector of probabilities $\boldsymbol{\eta}_i$ by a dichotomous vector. And we go from the MCLCM's to the PMD models by replacing every realization of $\boldsymbol{Z}$ by a vector of probabilities $\boldsymbol{\zeta}_p$.

MCLCM's are IRT models. They specify the relation between an observed item response and one or more latent variables characterizing the persons. Since every person is characterized by a vector, MCLCM's are related most to the multidimensional IRT models. There is a close relationship between the compensatory MCLCM in (8) and the item factor analysis model (Bock & Aitkin, 1981; Muthèn, 1978). In particular, the item factor analysis model is obtained by replacing the dichotomous Z_k's in (8) by continuous RV's.

There is also a close relationship between the conjunctive and the disjunctive MCLCM and the conjunctive and the disjunctive Rasch model (Embretson, 1980; Maris, 1995). The conjunctive and the disjunctive Rasch model are also LRM's. The difference with the corresponding MCLCM's is the PDF of the latent responses, the X_{ik}'s. For the MCLCM's, this PDF is given in (4). And for the two generalized Rasch models, this PDF is an ordinary Rasch model.

References

[1] Bishop, Y. M. M., Fienberg, S., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice.* Cambridge, Mass.: MIT Press.

[2] Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46,* 443-459.

[3] De Boeck, P., & Rosenberg, S. (1988). Hierarchical classes: Model and data analysis. *Psychometrika, 53,* 361-381.

[4] Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39,* 1-38.

[5] Embretson, S. (1980). Multicomponent latent trait models for ability tests. *Psychometrika, 45,* 479-494.

[6] Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian Data Analysis.* London: Chapman & Hall.

[7] Gill, P. E., Murray, W., & Wright, M. H. (1981). *Practical Optimization.* New York: Academic Press.

[8] 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.

[9] Hagenaars, J. A. (1990). *Categorical longitudinal data: loglinear panel, trend and cohort analysis.* Newbury Park, CA: Sage.

[10] Hagenaars, J. A. (1993). *Loglinear models with latent variables.* Newbury Park, CA: Sage.

[11] Heinen, T. (1993). *Discrete latent variable models.* Tilburg (The Netherlands): Tilburg University Press.

[12] Kelderman, H., & Rijkes, C. P. M. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika, 59*, 149-176.

[13] Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis.* Boston: Houghton Mifflin Company.

[14] Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B, 44*, 226-233.

[15] McLachlan, G. J., & Basford, K. E. (1988). *Mixture Models.* New York: Marcel Dekker.

[16] Maris, E. (1995). Psychometric latent response models. *Psychometrika, 60,* 523-547.

[17] Maris, E., De Boeck, P., & Van Mechelen, I. (1996). Probability matrix decomposition models. *Psychometrika, 61,* 7-29.

[18] Mood, A., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics.* Tokyo (Japan) : McGraw-Hill.

[19] Muthèn, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika, 43,* 551-560.

[20] Novick, M. R., & Jackson, J. E. (1974). *Statistical methods for educational and psychological research.* New York: McGraw-Hill.

[21] Rubin, D. B. (1987). A non-iterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest. Discussion of Tanner & Wong (1987). *Journal of the American Statistical Association, 82,* 543-546.

[22] Tatsuoka, K. K. (1984). *Analysis of fraction addition and subtraction problems.* (NIE Final Report). Urbana, IL: University of Illinois, Computer-based Education Research.

[23] Van Mechelen, I., De Boeck, P., & Rosenberg, S. (1995). The conjunctive model of hierarchical classes. *Psychometrika, 60,* 505-521.

Recent Measurement and Research Department Reports:

97-1 H.H.F.M. Verstralen. A Logistic Latent Class Model for Multiple Choice Items.

97-2 N.D. Verhelst. A New Heuristic for Estimating the Reliability of a Test.

97-3 N.D. Verhelst and R.J.H. Engelen. Estimating Latent Abilities from Raw Test Scores.

97-4 G.J.J.M. Straetmans and T.J.H.M. Eggen. Comparison of Test Administration Procedures for Placement Decisions in a Mathematics Course.

97-5 H.H.F.M. Verstralen. A Latent IRT Model for Options of Multiple Choice Items.

97-6 T.J.H.M. Eggen. On the Loss of Information in Conditional Maximum Likelihood Estimation of Item Parameters.

97-7 N.D. Verhelst and H.H.F.M. Verstralen. Modeling Sums of Binary Responses by the Partial Credit Model.