COMMUNICATIONS IN STATISTICS
Theory and Methods
Vol. 33, No. 6, pp. 1277–1294, 2004

Clustering Binary Variables in Subscales Using an Extended Rasch Model and Akaike Information Criterion

J.-B. Hardouin[1,*] and M. Mesbah[2]

[1]Regional Health Observatory, Hospital of Orléans,
Orléans, France
[2]UFR SSI, University of South Brittany,
Vannes, France

ABSTRACT

In this work, we handle the problem of selection of dichotomous items (questions with two possible answers) of a Quality of Life (QoL) questionnaire in sub-scales (subgroup of items producing unidimensional score). A procedure of clustering binary variables (items) in sub-scales with nice measurement properties is proposed. It is based on a new multidimensional Rasch model chosen in order to guarantee some specific measurement properties to the produced

*Correspondence: J.-B. Hardouin, Regional Health Observatory, Hospital of Orléans, 1, rue Porte Madeleine, BP 2439, 45032 Orléans Cedex 1, France; Fax: +33-238-744-881; E-mail: jean-benoit.hardouin@neuf.fr.

DOI: 10.1081/STA-120030149 0361-0926 (Print); 1532-415X (Online)

scores. The proposed process is presented, discussed and compared by simulations with the Mokken scale procedure (MSP). These simulations show that this new procedure is promising, specially when the structure of the set of binary variables is multidimensional, even if, several drawbacks persist, specially the time of computing of the procedure.

Key Words: Quality of life; Multidimensionality; IRT; Rasch model; AIC; Scales.

1. INTRODUCTION

As in the field of educational testing, in Health related QoL assessment, it is generally argued that QoL is simultaneously governed by several common latent traits: it is a multidimensional concept. But, even if existing instruments (questionnaires) are clearly divided in several subscales, the data are generally analyzed using separate and independent univariate models instead of a global multidimensional one.

In Item Response Theory (IRT) literature, there is plenty of unidimensional models proposed by various different authors. At the opposite side, published multidimensional models are rare and interpretation of their parameters is often not easy.

So, it will be useful to define a multidimensional IRT model easy to interpret. In this paper, we propose a new one: the Multidimensional Marginally Sufficient Rasch Model (MMSRM). We show how it is a satisfying multidimensional counterpart of the Rasch model with similar parameters easy to interpret.

Then, we present a new explanatory procedure, based mainly on the use of this model, to detect the relationships between the items and the latent traits.

Finally, with the helpful of this process, we show how to build sub-scales of items which must have, marginally, a good fit to Rasch model, and so, allow us to the dimensionality of the data. This process is compared by simulations to another existing empirical procedure: the Mokken scale procedure (MSP) based on the non parametric Mokken scales.

2. THE IRT MODELS

2.1. Notations

The IRT allow to analyse the responses from a sample of individuals to a bank of items, with subjective responses. Its scope of application is subjective or imprecise measurement (educational testing, psychology, sociology, etc...). We will consider in this paper only the specific field of QoL, which is a new and promising area of application.

QoL studies rely on questionnaires (psychometrical "tests" or instruments used to measure the latent traits) composed of J closed questions (items). In this paper, we consider only dichotomous items, but all the methods and ideas introduced here can be extended to polytomous case. The responses are coded 0 or 1: 1 is named "correct or positive answer" and 0 is the "uncorrected or negative answer".

For N individuals in a sample, the nth individual's response to item j is denoted by X_{nj}, for $n = 1, \ldots, N,\ j = 1, \ldots, J$.

Each subject is supposed as characterized by Q latents traits (*unobserved* aspects of Qol) and each one of them is represented by an unknown parameter Θ_q, $q = 1, \ldots, Q$ defined on $\mathbb{R}$: a low value of Θ_q represent a poor level in this quality of life continuum and vice versa.

This subject parameter can be assumed as a fixed or as a random parameter, i.e., an unobserved random variable. In the random case, the vector $\boldsymbol{\Theta}$ is defined as $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_q, \ldots, \Theta_Q)$, and each realization of this random vector is "attached" to an individual n and noted $\boldsymbol{\theta}_n = (\theta_{n1}, \ldots, \theta_{nq}, \ldots, \theta_{nQ})$.

The Item Response Function (IRF) of the item j is the function $P(X_{nj} = 1/\boldsymbol{\theta}_n; \boldsymbol{\nu}_j)$ where $\boldsymbol{\nu}_j$ is a set of parameters characterizing the item j.

2.2. Assumptions

The Item Response Theory rely on three fundamental assumptions:

- Fixed dimension which indicate that the dimension of the latent trait (Θ) is known: all the responses to the items are "governed" (produced) by a known number of common characteristics (the latent traits).
- Local independence which means that the answers of *one given individual* to two distinct items are independent. Such definition allow us to assume the latent as fixed as well as random parameter.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

When the latent (individual) parameter is assumed as random, local independence just means that the variables X_{nj_1} and X_{nj_2} with $j_1 \neq j_2$ are independent conditionally to $\boldsymbol{\theta}_n$. Its causal interpretation can be formulated as "all systematic covariation between items is explained by the latent trait(s) variation" (Molenaar, 1995).

- Monotonicity which means that the IRF are not decreasing in $\theta_{nq}, \forall q = 1, \ldots, Q$.

2.3. The Rasch Model

Generally, only unidimensional models are used in Qol Analysis, that is to say that the dimension is fixed to $Q = 1$. Various unidimensional models are used in practice, but the Rasch model (Rasch, 1960) is certainly the most famous one. It's the simplest model, and the easiest to interpret. In this model, each item is characterized by an unique parameter δ_j. The IRF of an item j is:

$$P(X_{nj} = 1/\theta_n; \delta_j) = \frac{\exp(\theta_n - \delta_j)}{1 + \exp(\theta_n - \delta_j)} \tag{1}$$

The δ_j parameter represent the difficulty of the item j: the more its value is high, the more the probability to correctly respond to this item is low for a given value of the latent trait. The Rasch model present a nice specific property: the total individual score $s_n = \sum_{j=1}^{J} x_{nj}$ is a sufficient statistic of the latent trait, or, in others words, the distribution of a response vector $\mathbf{X_n}$ conditionally to the observed score s_n is independent of the latent trait θ_n.

$$\begin{aligned} P(\mathbf{X_n} = \mathbf{x_n}/\theta_n; s_n, \delta_1, \ldots, \delta_J) &= \frac{P(\mathbf{X_n} = \mathbf{x_n}/\theta_n; \delta_1, \ldots, \delta_J)}{P(S_n = s_n/\theta_n; \delta_1, \ldots, \delta_J)} \\ &= P(\mathbf{X_n} = \mathbf{x_n}/s_n, \delta_1, \ldots, \delta_J) \end{aligned} \tag{2}$$

This result is a consequence of the factorization theorem applied to the exponential family. It follows directly when one rewrite the joint distribution of the items responses with the latent trait considered as unknown *fixed* parameters. When the latent trait is considered or assumed as random, the sufficiency of a statistics s_n for the random parameter θ_n is just a graphical property given by the same conditional independence property: the distribution of the response vector $\mathbf{X_n}$ conditionally to the observed score s_n is independent of the latent trait θ_n.

The Rasch model is the only IRT model with this property (Andersen, 1977). So, if we want to reduce all the information of the J items about the latent trait in only one statistic, we use the score S_n. On the other hand, the Rasch model is a very restrictive model and in practice, a good fit of the data to this model is difficult to achieve.

In statistical theory, the sufficiency property, added with the completness of the family distribution allows to get optimal estimate of the (latent) parameter among unbiased estimators. In our knowledge, in IRT, there is no available method of unbiased or natural estimators of the parameters. In practice, the sufficiency property and its simplicity is enough to justify the use of the total score as as surrogate of the latent trait: it is good enough. So, in IRT, estimators are often obtained through various maximum likelihood methods. They are generally biased, but consistency and asymptotical normality is often achieved.

Traditionally, when the latent is assumed as a fixed parameter, the difficulty parameters of the Rasch model are consistently estimated in maximizing the conditional likelihood obtained after conditioning over the observed total score of individuals (Andersen, 1970). When assuming the latent trait as a random variable, these parameters can be consistently estimated in maximizing the marginal likelihood, obtained after the integration of the joint distribution (of the items responses and of the latent variable) over this *unobserved* random variable (latent). Theoretically, any distribution can be specified. Nevertheless, generally, but not obligatory, in QoL psychometrical analysis, the distribution function of this latent variable is generally assumed as a centered Gaussian distribution with unknown variance σ^2 in QoL. Then, the marginal likelihood is:

$$L_M(\sigma^2, \delta_1, \ldots, \delta_J/\mathbf{x}) = \prod_{n=1}^{N} \int_{\mathbb{R}} \prod_{j=1}^{J} P(X_{nj} = x_{nj}/\theta; \delta_j) G(\theta/\sigma^2)\, d\theta \qquad (3)$$

2.4. Multidimensional Models

Kelderman and Rijkes (1994) propose an analytic multidimensional model based on a logistic form of the IRF of each item. The model permit to consider Q common latent traits influencing the responses of J items ($Q < J$). The response function of the jth item is given by:

$$P(X_{nj} = 1/\boldsymbol{\theta}_n; \delta_j) = \frac{\exp\left(\sum_{q=1}^{Q} B_{jq}\theta_{nq} - \delta_j\right)}{1 + \exp\left(\sum_{q=1}^{Q} B_{jq}\theta_{nq} - \delta_j\right)} \qquad (4)$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

The quantities B_{jq} are considered as *a priori* known fixed integers. The authors show that the vector of the sub-scores $\mathbf{s_n} = (s_{n1} \cdots s_{nq} \cdots s_{nQ})$, where s_{nq} are the realizations of the random variables $S_{nq} = \sum_{j=1}^{J} B_{jq} X_{nj}, \forall q$, is a sufficient statistic of the multidimensional latent trait $\boldsymbol{\theta}_n = (\theta_{n1} \cdots \theta_{nq} \cdots \theta_{nQ})$, that is to say that the distribution of a vector response $\mathbf{X_n}$ conditionally to the vector $\mathbf{s_n}$ is independent of the multidimensional latent trait $\boldsymbol{\theta}_n$.

As, in the unidimensional Rasch Model, the difficulty parameters δ_j can be consistently estimated by marginal maximum likelihood with the vector of latent traits assumed following a given multivariate distribution. For similar reasons as in unidimensional case, in QoL psychometrical analysis, specification of a multivariate centered Gaussian distribution with an unknown Σ variance matrix is generally preferred. Then the obtained marginal likelihood is:

$$L_M(\Sigma, \delta_1, \ldots, \delta_J / \mathbf{x}) = \prod_{n=1}^{N} \int_{\mathbb{R}^Q} \prod_{j=1}^{J} P(X_{nj} = x_{nj} / \boldsymbol{\theta}; \delta_j) G(\boldsymbol{\theta}/\Sigma) \, d\boldsymbol{\theta} \tag{5}$$

2.5. The Multidimensional Marginally Sufficient Rasch Model (MMSRM)

Among unidimensional IRT models, the Rasch model is the most famous, mostly because the property of sufficiency of the individual total score for the latent trait make this model in demand. So, when someone use the individual total score to summarize the responses of items, he assumes, knowingly or not, that the model governing his data is a Rasch model. In the multidimensional case, the notion of sufficiency of the (multivariate) score on the (multivariate) latent trait is not trivial to define. The Kelderman model is a model where the vector of sub-scores is a sufficient statistics for the vector of latent traits: it is a "global" sufficiency property. A strong definition of sufficiency, that we call "marginal" sufficiency, is that each sub-score must be sufficient for the corresponding latent trait. The new multidimensional family model that we propose, in the following, verify this property: each sub-scale follow a Rasch model and consequently, each sub-score s_{nq} is a sufficient statistic of the corresponding latent trait θ_{nq}.

We name such family model: Multidimensional Marginally Sufficient Rasch Model (MMSRM).

A model is constructed in using Q distinct sets of items, with each set verifying an unidimensional Rasch model. The multidimensional latent

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

trait is formed by the union of the Q scalar latent traits. The IRF of the jth item of the sub-scale X^q is given by:

$$P\left(X_{nj}^q = 1/\boldsymbol{\theta}_{nq}; \delta_j^{(q)}\right) = \frac{\exp\left(\theta_{nq} - \delta_j^{(q)}\right)}{1 + \exp\left(\theta_{nq} - \delta_j^{(q)}\right)} \tag{6}$$

Here X_{nj}^q is the observed response of subject n to item j belonging to the sub-scale q. The parameters $\delta_j^{(q)}$ correspond to the difficulty of the jth item with respect to the latent trait which this item is rely.

This is not enough to specify our model; we need to add the conditional independence properties which will guarantee to us the separation of all marginal models.

First, we build a conditional model, i.e., we specify only the conditional distribution $f(X/\boldsymbol{\theta})$ (distribution of the responses to items conditional to the latent): How? This global conditional distribution is specified by *construction*, as the product of the conditional distributions $f_q(X^q/\theta_q)$ of the Q sub scales. Each of these conditional model is chosen by *construction* as an unidimensional Rasch Model.

$$f(X/\boldsymbol{\theta}) = \prod_{q=1}^{Q} f_q(X^q/\theta_q) \tag{7}$$

Then, we specify a multidimensional joint distribution $G(\boldsymbol{\theta})$, for the multivariate vector of latent $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_Q)$. We choose for convenience and easiness of interpretation this joint distribution as centered multivariate gaussian.

Then, the joint distribution of the items and the latent is given by:

$$f(X, \boldsymbol{\theta}) = \prod_{q=1}^{Q} f_q(X^q/\theta_q) G(\boldsymbol{\theta}) \tag{8}$$

Then, it will be straightforward to show, that, in the global conditional model, in marginalizing the joint distribution over all other items not included in a given sub-scale, we get a Rasch model, and then its sufficiency property. The differences between a Kelderman model and a MMSRM are illustrated by the Fig. 1 who present graphical representations of these two models, and the principle of sufficiency of the sub-scores on the latent traits in the two models.

The parameters $\delta_j^{(q)}$, $\forall j = 1, \ldots, J$ and the variance–covariance matrix of the latent traits Σ can be consistently estimated by the marginal maximum likelihood in maximizing a quantity similar to one defined in (5).

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

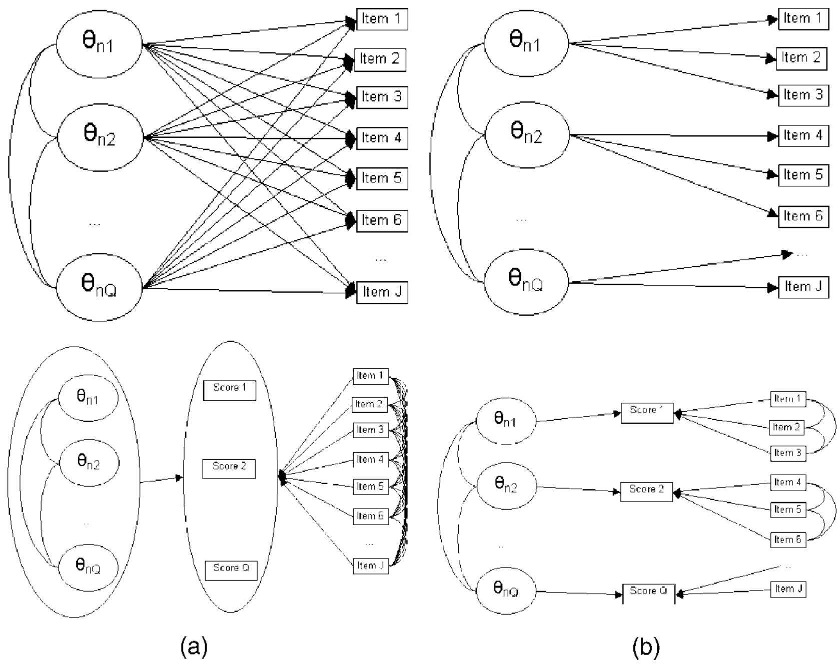

Figure 1. (a) The Kelderman model. (b) The MMSRM. The Kelderman model and the MMSRM and the principle of sufficiency of the sub-scores on the latent traits under these two models.

3. THE SELECTION OF ITEMS IN SUBSCALES WITH THE MMSRM

3.1. Aim

In IRT, practitioners usually search to find the relationships between items and dimensions: in others words, they search to build unidimensional sub-scales. Indeed, instruments are generally subjectively defined by psychologists, i.e., using only their a priori knowledge. Then, IRT models are used only to confirm, using real data, some a priori assumed relationships between the items and the latent traits: those models are used in a confirmatory but rarely in an exploratory kind of analysis.

More and more, psychologists or practitioners, want to build instruments objectively. They produce items using focus groups or available literature on the subject, then they search to found the various dimensions of the instruments and relationships between items and dimensions, using real data. This step, of course, can also include an a priori knowledge. Most often, they use factorial methods but the qualitative (binary or ordinal) nature of the items is a complication: factorial methods deals mainly with quantitative data. More, the interpretation of the results of factorial methods is not always obvious and the choices of practitioners can be questionable. So, often, empirical results show that sub-scales obtained by factor analysis don't fit well to unidimensional IRT models, and in particular to the Rasch model.

Several exploratory methods have been proposed to classify the items in different dimensions. Some of them are based on factor analysis as the Revised Modified Parallel Analysis (Budescu et al., 1997) and others on the properties of the IRT like the Mokken scale procedure (MSP) (Hemker et al., 1995), the DETECT method (Zhang and Stout, 1999) or the hierarchical cluster analysis with conditional proximity measures (HCA/CCPROX) (Roussos et al., 1998) but none of them allow to obtain automatically sub-scales constrained to fit a Rasch model. We propose a method based on the fit of the data to a MMSRM, because each sub-scale of a MMSRM verify a Rasch model. We explain, in this part, the principles of this method, and we present a set of various simulations with a comparison between this method and the Mokken Scale Procedure.

3.2. Principles and Algorithm

We suppose that the structure of the items is simple, that is to say that each item is related to only one latent trait (the response of one item depend only of the value of the latent trait to which this item is related). The number of latent traits is unknown. We want to find the partition of items among all possible partitions who give the best fit to a MMSRM with all the items. The fit of a model is valuated by the Akaike Information Criterion (*AIC*): the model with the lower *AIC* is the more parsimonious model, that is to say that this model explain the more important amount of information compared to its number of parameters. We note $l_{max}(\boldsymbol{\beta}/\mathbf{x})$ the value of the maximal log-likelihood of a N-sample obtained when the chosen model include a set of parameter, noted $\boldsymbol{\beta}$, a vector of dimension k. The Akaike criterion associated to this model and the N-sample is:

$$AIC = -2l_{max}(\boldsymbol{\beta}/\mathbf{x}) + 2k \tag{9}$$

We can used the maximum value of a marginal log-likelihood as ones defined in Eq. (5) as $l_{max}(\boldsymbol{\beta}/\mathbf{x})$. In a MMSRM, the $\boldsymbol{\beta}$ vector is composed of the parameters of the distribution of the latent trait (the $(Q^2 + Q)/2$

components of the Σ matrix) and of the J difficulty parameters of the items $\delta_j^{(q)}$, so $k = J + (Q^2 + Q)/2$.

In practice, it is not possible to compute the AIC of all the possible partitions of items in a reasonable time, because the estimation of parameters of the model takes a long time, even with very fast computers. So, we propose an iterative relatively fast process which permit first, to define the number of latent traits, and second to find the links between the items and the latent traits.

The main idea of the method is to compare the fit of the data concerning the responses of an one dimensional set of items in adding a new item in the same dimension or in an other dimension. This method is based on the forward principle. To construct each sub-scale, we use the following iterative process:

- *At the initial step*, we consider two items supposed to be unidimensional: we choose the pair of items (j_1, j_2) which have the higher coefficient $H_{j_1 j_2}$ of Loevinger (1948) among all the possible pairs of items. This coefficient is used because it permit to rapidly measure the amount of unidimensionality between two items.

  The others items are classified, using the Loevinger coefficient H_j^S (1948), from the item who probably own the less to the same dimension as the initial pair (first place) to the item who probably own the more to this dimension (last place).

  The coefficients $H_{j_1 j_2}$ and H_j^S are defined in the Appendix.
- *At the step* k, the item classified at the kth place is selected in the current subscale, if the AIC of the unidimensional Rasch model including all the already selected items and the kth item is smaller than the AIC of the MMSRM built with all the selected items in a first dimension and the kth item in a second dimension.

When the first subscale is obtained, we repeat the same process with all the remaining items, and so on, until there is no more items remaining.

To allow a gain of computing time, the subscales built in a previous step are excluded from the goodness of fit testing process when we are building a new subscale.

3.3. Parameters of the Simulations

The proposed procedure had been empirically tested with simulated data. We always have considered two dimensions in the simulations but we suppose similar results with higher dimensional data. The parameters used in the simulations are similar to ones used in a comparative study between the procedures MSP, DETECT and HCA/CCPROX (Van Abswoude, 2001):

- The used model to simulate data: multidimensional extensions of the 2-PLM (two-parameters logistic model) and of the 5-PAM (five-parameters accelerated model) – see below.
- The structure of the items (two kinds of structures: simple structure or approximate simple structure – see below).
- The correlation coefficient ρ between the two simulated latent traits (six levels: 0.0, 0.2, 0.4, 0.6, 0.8 and 1.0).
- The strength of the relation between the items of each dimension and the latent trait which this dimension is the more strongly related, measured by a α_q parameter (discriminating power) (three levels: low (0.4), medium (0.7) and high (1.7)).
- The number of items in each dimension (two main cases: 7 items per dimension or 7 items in one dimension and 14 in the other).

In a simple structure, the response variable to each item is influenced by only one latent trait. In this case, the used extensions of the 2-PLM is:

$$P\left(X_{nj}^q = 1/\boldsymbol{\theta}_n; \alpha_q, \delta_j^{(q)}\right) = \frac{exp\left[1.7\left(\alpha_q \theta_{nq} - \delta_j^{(q)}\right)\right]}{1 + exp\left[1.7\left(\alpha_q \theta_{nq} - \delta_j^{(q)}\right)\right]} \tag{10}$$

and this one of the 5-PAM is :

$$\begin{aligned} P(X_{nj}^q &= 1/\boldsymbol{\theta}_n; \alpha_q^*, \delta_j^{*(q)}, \gamma_j^{low}, \gamma_j^{up}, \xi_j) \\ &= \gamma_j^{low} + (\gamma_j^{up} - \gamma_j^{low}) \left(\frac{exp\left[1.7\left(\alpha_q^* \theta_{nq} - \delta_j^{*(q)}\right)\right]}{1 + exp\left[1.7\left(\alpha_q^* \theta_{nq} - \delta_j^{*(q)}\right)\right]} \right)^{\xi_j} \end{aligned} \tag{11}$$

with, for each item, the three parameters γ_j^{low}, γ_j^{up} and ξ_j, respectively, fixed to 0.1, 0.9 and 2. The coefficient 1.7 is a very classical coefficient used in IRT (Van Abswoude, 2001). In an approximate simple structure, we consider that the response to an item j of the sub-scale X^q is strongly affect by θ_q and weakly by $\theta_{\bar{q}}$ where $\bar{q}$ is the complementary of q in $\{1; 2\}$ (we used a coefficient of $\alpha_{\bar{q}} = 0.2$ to weight this weak link). Equations (10) and (11), respectively, become:

$$P(X_{nj}^q = 1/\boldsymbol{\theta}_n; \alpha_q, \alpha_{\bar{q}}, \delta_j^{(q)}) = \frac{exp\left[1.7\left(\alpha_q \theta_{nq} + \alpha_{\bar{q}} \theta_{n\bar{q}} - \delta_j^{(q)}\right)\right]}{1 + exp\left[1.7\left(\alpha_q \theta_{nq} + \alpha_{\bar{q}} \theta_{n\bar{q}} - \delta_j^{(q)}\right)\right]} \tag{12}$$

and this one of the 5-PAM is :

$$P(X_{nj}^q = 1/\boldsymbol{\theta}_n; \alpha_q^*, \alpha_{\bar{q}}^*, \delta_j^{*(q)}, \gamma_j^{low}, \gamma_j^{up}, \xi_j)$$
$$= \gamma_j^{low} + (\gamma_j^{up} - \gamma_j^{low}) \left(\frac{exp\left[1.7\left(\alpha_q^*\theta_{nq} + \alpha_{\bar{q}}^*\theta_{n\bar{q}} - \delta_j^{*(q)}\right)\right]}{1 + exp\left[1.7\left(\alpha_q^*\theta_{nq} + \alpha_{\bar{q}}^*\theta_{n\bar{q}} - \delta_j^{*(q)}\right)\right]} \right)^{\xi_j} \tag{13}$$

In Eqs. (11) and (13), the parameters α_q^*, $\alpha_{\bar{q}}^*$ and $\delta_j^{*(q)}$ are computed to obtain IRF with the same maximal slope and the same location of this maximal slope than the IRF present in Eqs. (10) and (12) (Van Abswoude, 2001).

We easily prove than the model presented in Eq. (10) is a MMSRM in using the random variable $\tilde{\theta}_{nq} = 1.7\alpha_q\theta_{nq}$ and the parameters $\tilde{\delta}_j^{(q)} = 1.7\delta_j^{(q)}$. Equation (10) become:

$$P\left(X_{nj}^q = 1/\tilde{\boldsymbol{\theta}}_n; \tilde{\delta}_j^{(q)}\right) = \frac{exp\left[\tilde{\theta}_{nq} - \tilde{\delta}_j^{(q)}\right]}{1 + exp\left[\tilde{\theta}_{nq} - \tilde{\delta}_j^{(q)}\right]} \tag{14}$$

The latent traits are simulated with a standardized multivariate normal distribution. The difficulty parameters δ_j are chosen in each dimension as the percentiles of the standardized normal distribution (for the jth item of a given dimension q including J_q items, $\delta_j^{(q)} = z^{-1}\left(j/(J_q + 1)\right)$ where $z(x)$ is a standardized normal distribution function). Each simulation concern 2000 of individuals. All the 360 possible combinations of the parameters used in the simulations have been used one time.

As a consequence, in Eq. (14), the latent trait $\tilde{\boldsymbol{\theta}}$ is simulated with a centered multivariate Gaussian distribution with a variance matrix

$$\Sigma = \begin{pmatrix} (1.7\alpha_1)^2 & 1.7^2\rho\alpha_1\alpha_2 \\ 1.7^2\rho\alpha_1\alpha_2 & (1.7\alpha_2)^2 \end{pmatrix}$$

As the MMSRM is a Generalized Linear Mixed Model (GLMM) with a logistic function as link function and Q dependent random variables, the parameters $\delta_j^{(q)}$ and Σ can be consistently estimated by adapted

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

procedure of classical softwares as NLMIXED procedure for SAS software or GLLAMM for Stata software.

Indeed, the jointly estimation of the items parameters $\delta_j^{(q)}$ and of the variance matrix Σ is a very long process, so, to gain a considerably amount of computing time, the parameters of the model are estimated in two steps: at a first step, the marginal maximum likelihood estimators of the difficulty parameters are obtained in each subscale in fitting the data to an unidimensional Rasch model (the XTLOGIT procedure of Stata is used); at a second time, these estimators are used in the multidimensional model and the parameters of the distribution of the multidimensional latent trait are estimated by marginal maximum likelihood (with the GLLAMM procedure of Stata, in approximating the multidimensional integrals by adaptive Gaussian quadratures and in using a quasi-Newton algorithm). It turned out that this method of estimation, previously used by others in practice (Kelderman and Rijkes, 1994), don't strongly affect the estimations of the parameters.

3.4. Classification of the Results

Each simulated database is submitted to the proposed procedure and to MSP. The results of MSP depend of an arbitrary fixed threshold c. Three different values for the threshold c are used: 0.3, 0.2 and 0.1 (Hemker et al., 1995). The first value of the threshold is one proposed by the authors of MSP, but this value don't always permit to select all the items in the different subscales. The more this threshold is low, the more the items are easily selected in the subscales, and the more the results can be analysed from a determinate way. So the two last values are been chosen in this aim, even if too low value for this threshold product full of classification errors.

The simulations are classified in 7 groups, following their results, as described in Table 1.

3.5. Analysis of the Results of the Simulations

The model used to simulate the date and the correlation between the two original latent traits play an important role in the results of the procedure. When we simulate the data with a MMSRM (simple structure and extension of the 2-PLM, see Eq. (14)), the results obtained with our procedure are very good, except when the two simulated latent traits are confused (when $\rho = 1$).

Table 1. Classification of the results of the simulation.

Class	Interpretation
1	The true classification of the items is found.
2	Only one item is not classified in the two first dimensions (minor error).
3	The true classification is not found but there is no pair(s) of items who come from the two distinct simulated dimensions who are classified together.
4	There is one or more pair(s) of items who come from the two distinct simulated dimensions who are classified together.
5	All the items are classified in only one dimension (it is a good results when the correlation between the two dimensions is 1.0).
Specific results for MSP	
U	All the selected items are classified in only one dimension but all the selected items come from the same simulated dimension.
N	No item have been selected.

Note: With MSP, a result of type 3, 4 or 5 can represent an incomplete partition of items.

Concerning the influence of the correlation between the two simulated latent trait, we can separate the results in three classes (Table 2):

- When the correlation between the two simulated latent traits is low (inferior or equal to 0.4), the proposed procedure present the best rate of correct results (type 1 or 2) compared with MSP whatever the used value of the threshold c and a correct rate of bad results (similar to MSP with c fixed to 0.2).
- When the correlation between the two simulated latent traits is high (superior or equal to 0.6), the proposed procedure present the best rate of correct results and a rate of bad results similar or inferior to MSP.
- When there is two confused simulated latent traits ($\rho = 1$), the procedure is not efficient because it search to distinguish in several subscales the items differently rely to the only one latent trait (it is the case when the items of the two dimensions are simulated with different discriminating power: $\alpha_1 \neq \alpha_2$). Indeed, it is a logical result in the Rasch model where a good fit is realized only for items with a similar discriminating power.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Table 2. Global results of the simulations in function of the used procedure and of the correlation between the two simulated latent traits.

Used model		MMRSM			Others models		
Procedure	Results	$\rho \leq 0.4$	$0.6 \leq \rho \leq 0.8$	$\rho = 1.0$	$\rho \leq 0.4$	$0.6 \leq \rho \leq 0.8$	$\rho = 1.0$
Number of simulations		45	30	15	135	90	45
Proposed	Good	34(75%)	19(38%)	1(8%)	61(45%)	26(29%)	4(9%)
	Bad	5(11%)	9(30%)	7(48%)	48(36%)	53(59%)	8(18%)
MSP	Good	28(62%)	8(27%)	8(53%)	38(28%)	8(9%)	12(27%)
($c = 0.3$)	Bad	0(0%)	15(50%)	2(13%)	27(20%)	48(53%)	3(25%)
	Unspecified	4	1	2	39	20	9
MSP	Good	18(40%)	2(7%)	11(73%)	48(36%)	7(8%)	38(84%)
($c = 0.2$)	Bad	6(13%)	21(70%)	0(0%)	57(42%)	70(78%)	1(2%)
	Unspecified	1	0	1	18	11	1
MSP	Good	25(56%)	0(0%)	14(93%)	20(15%)	1(1%)	43(96%)
($c = 0.1$)	Bad	17(38%)	30(100%)	0(0%)	104(77%)	87(97%)	0(0%)
	Unspecified	1	0	1	1	0	0

Logically, we note, for each procedure, better results when the discriminating power of the items are strong than when they are weak. In the case where the data are simulated by a MMSRM, this phenomenon can be interpret as the fact that the procedure give better results when the latent traits have large dispersions, than when these dispersions are small.

CONCLUSIONS

The new procedure seems to be more efficient to class correctly all the items than MSP, whether the data are simulated with a MMSRM, or not. Compared to MSP, the proposed procedure had the advantages not to depend to a threshold fixed arbitrary by the user, to detect more often the structure of items than MSP with a rate of errors similar than MSP with a high value of c, and to not give unspecified results.

In the case where items are relied with different strength to the latent trait, the proposed procedure classify together items with similar discriminating powers: this phenomenon is due to the fact that, in the used model, in this procedure, each sub-scale verify a Rasch model who suppose equal discriminating powers of all the items.

Globally, the new procedure allow to build sub-scales verifying a Rasch model. The results are encouraging, but some drawbacks must be reduced:

- The time of computing is the main drawback (the process with 7 items in each dimensions take about 24 h with a computer cadenced at 120 MHz and 32 Mo of RAM or about 7 h with a computer cadenced at 950 MHz and 512 Mo of RAM) and it could be interesting to study others estimations techniques as Generalized Estimating Equations (GEE) or conditional maximum likelihood to improve it.
- The rate of bad results (36% on all the simulations) could certainly be reduced in adapting the algorithm, for example in using a stepwise instead of a forward procedure. With a stepwise procedure, we remove, at each step, previously included items which have a bad fit at the actual step (actually, only the global fit to a Rasch model is valuated by the Akaike criterion, but the fit of each item is never studied).
- The procedure must be extended to polytomous items, a kind of items very current in quality of life studies. That needs to define, when items are polytomous, a multidimensional family of models similar to the MMSRM family.

APPENDIX: THE LOEVINGER COEFFICIENTS

In the proposed procedure, we use the Loevinger coefficient $H_{j_1 j_2}$. This index is defined by Loevinger (1948) and is an index of unidimensionality. We put N the number of individuals, and X_j the random variable who correspond to the response to the item j (with 2 possible responses: 0 (negative response) and 1 (positive response)).

For two item j_1 and j_2 with $P(X_{j_1} = 1) < P(X_{j_2} = 1)$, we define the quantities:

$$e_{j_1 j_2} = NP\big(X_{j_1} = 1, X_{j_2} = 0\big) \quad \text{and} \quad e^{(0)}_{j_1 j_2} = NP(X_{j_1} = 1)P(X_{j_2} = 0)$$

The Loevinger coefficient between these two items is equal to:

$$H_{j_1 j_2} = 1 - \frac{e_{j_1 j_2}}{e^{(0)}_{j_1 j_2}}$$

The Loevinger coefficient $H_{j_1 j_2}$ take a value of 1 if none individuals had correctly respond to the more difficult item, and negatively respond to the easier item. In this case, the two item measure exactly the same concept, so this pair of items is unidimensional. It take a value near 0 if the responses to the two items are independent. In this case, the two items don't measure the same concept and are not unidimensional.

Loevinger too define a coefficient of integration of one item in a scale. We note S the set of indexes of J items composing a scale: $S = \{j_1, j_2, \ldots, j_J\}$. The Loevinger coefficient of the item k, $k \notin S$ in this scale is equal to:

$$H_k^S = 1 - \frac{\sum_{j \in S} e_{jk}}{\sum_{j \in S} e^{(0)}_{jk}} = \frac{\sum_{j \in S} e^{(0)}_{jk} H_{jk}}{\sum_{j \in S} e^{(0)}_{jk}}$$

The Loevinger coefficient H_k^S take a value of 1 if all the items of the scale S and the item k exactly measure the same concept. If the scale S is unidimensional, a value of the coefficient H_k near 0 means that the item k don't measure the same concept as the scale S.

The Mokken Scale Procedure (MSP) (Hemker et al., 1995) permit to construct sub-scales verifying for each pair of items $(j_1; j_2)$ of each sub-scale $H_{j_1 j_2} > 0$ and for each items j of the sub-scale S $H_j^S > c$ where c is a threshold chosen by the user (the value $c = 0.3$ is proposed by the authors).

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

REFERENCES

Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *J. R. Statist. Soc. Series B.* 32:283–301.

Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika* 42(1):69–81.

Budescu, D. V., Cohen, Y., Ben-Simon, A. (1997). A revised modified parallel analysis for the construction of unidimensional item pools. *Appl. Psychol. Meas.* 21(3):233–252.

Hemker, B. T., Sijtsma, K., Molenaar, I. W. (1995). Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken IRT model. *Appl. Psychol. Meas.* 19:337–352.

Kelderman, H., Rijkes, C. P. M. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika* 59:149–176.

Loevinger, J. (1948). The technique of homogeneous tests compared with some aspects of "scale analysis" and factor analysis. *Psychol. Bull.* 45:507–530.

Molenaar, I. W. (1995). Some background for Item Response Theory and the Rasch model. In: Fisher, G. H., Molenaar, I. W., eds. *Rasch Models. Foundations, Recent Developments, and Applications*. New York: Springer, 436 pp.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Nielsen & Lydiche.

Roussos, L. A., Stout, W. F., Marden, J. I. (1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *J. Educ. Meas.* 35:1–30.

van Abswoude, A. A. H., van den Ark, L. A., Sijtsma, K. (2001). A comparative study on test dimensionality assessment procedures under non parametric IRT models. Technical report. University of Tilburg.

Zhang, J., Stout, W. F. (1999). The theorical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika* 64:213–249.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016