Methods of Psychological Research Online 1999, Vol. 4, No. 3
Internet: http://www.mpr-online.de/

Collateral information and Mixed Rasch models

Arnold Smit[1,2]
Henk Kelderman[1]
Henk van der Flier[1]

Abstract

A simulation study is conducted to evaluate the usefulness of incorporating collateral information in the Mixed Rasch model. The results show that the standard errors as well as latent class membership assignment can benefit substantially from incorporating external variables that associate with the latent class variable. Especially when the difference in probability structure between the latent classes becomes smaller, or, when the sample size is relatively small.

Keywords : mixed Rasch model; item response theory; latent class analyses

[1]Department of Work & Organizational Psychology, Free University Amsterdam; v.d. Boechorststraat 1; 1081 BT, Amsterdam
[2]Netherlands Organization of Scientific Research (NWO) is gratefully acknowledged for funding this project. This research was supported by a grant of the Foundation for Behavioral and Educational Sciences of this organization (575-30-004).

1 Introduction

Latent variable models are, slowly but surely, finding their way to applied fields. These models thrived in the sociometric and psychometric community for decades, but, to a large extend, failed to reach more applied researchers. With the availability of user friendly software in combination with a large body of literature, ranging from non-technical guidelines to the more abstract mathematical foundations of the models, more researchers are considering the latent variable modeling framework. For an overview see Heinen (1993), Fisher and Molenaar (1995), Linden and Hambleton (1997), Rost and Langeheine (1997), Wilson, Engelhard, and Draney (1997), and Wilson, Draney, and Engelhard (1999).

The present paper is concerned with the Mixed Rasch Model. This hybrid model for analyzing dichotomous data, due to the work of Rost (1990, 1991), contains both latent trait and latent class variables. The Rasch model is often too strict for social science data. If the data cannot be summarized well enough (obtain a reasonable model fit) using the Rasch model, roughly two strategies can be adopted. First, more 'complex' IRT models could be fit. Second, one could try to disentangle the sample into Rasch scalable subgroups. We prefer the second approach from a measurement theoretical viewpoint. The Rasch model can be build from assumptions that may reasonably be demanded of scientific measurement. Furthermore, finding Rasch scalable subgroups gives substantial insight in the data. Interpreting differences in discrimination parameters or incorporating extra latent trait variables into the analysis is a rather daunting task if no a priori justification can be given. As an aside, if substantive knowledge is available about the actual cognitive process involved, inference can be stretched even further with a hybrid variant of the linear logistic test model (Fisher, 1973), see Mislevy and Verhelst (1990).

A problem with the Mixed Rasch model is that differentiation between latent subgroups can only be done successfully if the differences are sufficiently large. Here we evaluate the gain obtained by using external variables in identifying the latent discrete mixing variable. The main goal is to obtain more stable solutions, or, stated differently, to track more subtle effects of mixing variables. The ability to track subtle effects is especially important when dealing with a discrete latent bias variable. A general framework for modeling/assessing DIF is given in Kelderman and Macready (1990).

As an example, consider an exam where some questions were discussed during the last lecture. Students who attended that lecture will have a higher probability of giving the correct response. Furthermore, suppose we do not know who attended the last lecture, we only know that eighty percent of the female students and thirty percent of the male students attended the class. The question addressed here is if such extra (collateral) information can be exploited in the Mixed Rasch model.

Within the latent class analysis framework, the use of extra (typically called concomitant) variables was applied by Clogg and Goodman (1984), see also Dayton and Macready (1988), Gupta and Chintagunta (1994), and Böckenholt (1997). Within the item response theory framework Mislevy and Sheehan (1989a, 1989b) showed that, using the *missing information principle* of Orchard and Woodbury (1972), the incorporation of collateral variables (related to the latent trait) can reduce the standard errors of the parameter estimates. In what follows the general idea to use collateral variables in identifying the latent mixing variable and reducing standard errors is applied to the mixed Rasch model.

2 The Loglinear Mixed Rasch model

Let the random variable X_{ij}, taking values $\{0,1\}$, denote the response of person i to item j. To make things concrete, assume $X_{ij}=1$ if the correct answer is given and $X_{ij}=0$ if the wrong answer is given. Furthermore, let θ_{im} denote the latent trait, and δ_{jm} denote the item difficulty (or easiness) for members of latent class m.

We assume that the probability of a correct response, on item j by a person i who is member of latent class m, follows the Rasch model.

$$P_{ij|m} \equiv P\left(X_{ij}=1 \mid \theta_{im},\delta_{jm}\right) = \frac{e^{\theta_{im}+\delta_{jm}}}{1+e^{\theta_{im}+\delta_{jm}}} \tag{1}$$

The probability of the response pattern of person i, assuming local stochastic independence, is

$$P_{i|m} \quad = \quad \prod_{j=1}^{n} P_{ij|m} = \frac{e^{\sum_j x_{ij}(\theta_{im}+\delta_{jm})}}{\prod_j\left(1+e^{\theta_{im}+\delta_{jm}}\right)} = \frac{e^{t_i\theta_{im}+\sum_j x_{ij}\delta_{jm}}}{\prod_j\left(1+e^{\theta_{im}+\delta_{jm}}\right)} \tag{2}$$

where $t_i=\sum_j x_{ij}$ denotes the number of correct responses, or sum score. From the above formula we see that t_i is sufficient for θ_{im} within a latent class. In other words, the (nuisance) parameter θ_{im} can be eliminated by conditioning on t_i in latent class m.

More explicitly, let $S_t=\{\boldsymbol{x}:\sum_j x_j=t\}$ denote the set of response patterns with sum score t. The probability of a response pattern in this set for person i who belongs to class m, is

$$\begin{aligned} P_{t|m} \quad &= \quad \sum_{\boldsymbol{x}\in S_t} P_{i|m} = \frac{e^{t\theta_{im}}}{\prod_j\left(1+e^{\theta_{im}+\delta_{jm}}\right)} \sum_{\boldsymbol{x}\in S_t} e^{\sum_j x_{ij}\delta_{jm}} \\ &= \quad \frac{e^{t\theta_{im}}}{\prod_j\left(1+e^{\theta_{im}+\delta_{jm}}\right)}\gamma_t\left(\boldsymbol{\delta}_m\right) \end{aligned}$$

where $\gamma_t\left(\boldsymbol{\delta}_m\right)$ denotes the symmetric basis function, with the class specific vector of item difficulty parameters ($\boldsymbol{\delta}_m=\delta_{m1},\cdots,\delta_{mn}$) as argument. Now by conditioning on t and m we get

$$P_{i|tm} \quad = \quad \frac{P_{itm}}{P_{tm}} = \frac{P_{im}}{P_{tm}} = \frac{P_{i|m}}{P_{t|m}} = \frac{\prod_j e^{x_{ij}\delta_{jm}}}{\gamma_t\left(\boldsymbol{\delta}_m\right)}$$

Note that this expression is independent of θ_{im}. Thus a more general notation can be adopted where the person index i is replaced by an index denoting a response pattern, say ν. Furthermore, if the latent class is dichotomous, taking values $\{0,1\}$, then δ_{jm} can be reparameterized as $\delta_j + m\Delta_j$. Then, the log of the marginal probability is

$$\begin{aligned} \log P_{\nu m} &= \log P_{\nu|tm} P_{tm} \\ &= \log \frac{P_{tm}}{\gamma_t\left(\boldsymbol{\delta}_m\right)} + \sum_j \left(x_{\nu j}\delta_j + x_{\nu j} m \Delta_j\right) \end{aligned} \tag{3}$$

The response pattern *frequencies* have expected values given by the model

$$\log f_{\nu m} = \mu + \mu_t^T + \mu_m^M + \mu_{tm}^{TM} + \sum_{j=1}^{n} \left(\mu_{x_j}^{X_j} + \mu_{x_j m}^{X_j M}\right) \tag{4}$$

where $f_{\nu m}$ denotes the expected frequencies of response pattern ν in latent class m. the main parameters $\mu_{x_j}^{X_J}$ are the item difficulty parameters, and the interaction parameters $\mu_{x_j m}^{X_J M}$ are the differences in item difficulties between the latent classes.

We cannot fit a quasi independence model (the items are independent within the sum score $\times$ latent class cells), because class membership is not observed. Note that if m would be observed we simply have a Rasch Model with an observed grouping variable. Unobserved variables can be handled with the *EM*–algorithm. The idea is to make an initial guess of the complete unobserved table (marginals), estimate the parameters using this table, and with these parameters we can construct a new complete table (expected table given the current parameter estimates and the observed incomplete table). We repeat this procedure until differences become sufficiently small. The algorithm thus splits the observed table into unobserved subtables that are most likely given the model. It is clear that the amount of difference in probability structure of these subtables is related to the ability of the algorithm to disentangle the observed tables in meaningful subtables. Experimenting with these models reveals that if differences in probability structure become 'too' small the algorithm is splitting up the table to incorporate a few outliers. Typically, in these cases, solutions are obtained with very small class sizes and extreme parameter estimates within these classes. These solutions are no more than capitalization on chance.

3 Collateral information

Without loss of generality, assume we have one dichotomous collateral variable, say g, that is associated with the latent class variable. Furthermore, assume that the items are independent conditionally on the latent class and the sum score. This implies that g is redundant in describing the probability structure of the response patterns once we conditioned on the latent class and sum score. So the following decomposition is possible

$$P_{\nu mg} = P_{\nu|tmg} P_{tmg} = P_{\nu|tm} P_{tmg}$$

The final step in the specification of the model is to give a sensible model structure for the joint distribution P_{tmg}. The crucial observation is that the sum score should be independent of the collateral variable conditional on the latent class, otherwise the collateral variable is itself a bias variable. The association between them is used to identify the latent class variable by forcing independence conditional on the latent class variable. The probability can thus be factored further

$$\begin{aligned} P_{\nu mg} &= P_{\nu|tm} P_{t|m} P_{g|m} P_m \\ &= \left(P_{\nu|tm} P_{t|m}\right)\left(P_{m|g} P_g\right) \end{aligned}$$

And, the response pattern *frequencies*, written in the usual loglinear notation

$$\log f_{\nu mg} = \mu + \mu_t^T + \mu_m^M + \mu_g^G + \mu_{tm}^{TM} + \mu_{gm}^{GM} + \sum_{j=1}^{n}\left(\mu_{x_j}^{X_j} + \mu_{x_j m}^{X_j M}\right) \tag{5}$$

A simulation study is conducted to evaluate the use of collateral information in fitting the Mixed Rasch Model. The expectations are twofold. First, the variation of the parameter estimates will be smaller. The missing information principle (Orchard & Woodbury, 1972; Little & Rubin, 1987) decomposes the complete information in observed information and missing information. By incorporating a variable that associates with the missing variable (latent class), the missing information is expected to become smaller, which causes the observed information to become greater, and thus, the variability of the parameter estimates to shrink. Second, the extra information, related to the latent class, will render the algorithm *less* susceptible for converging to local solutions.

4 Simulation design

Data sets with ten items are generated according to the Mixed Rasch model, with a dichotomous latent class variable (equal class sizes). The number of subjects and the item parameter sets are varied, each with two levels. Data sets are generated with four hundred and with two thousand subjects (with $\theta \sim N[0,1]$), using one of the parameter sets in Table 1. In the sequel, the parameter sets are rather cryptically called 2TO6 and EVEN. As can be seen in table 1, 2TO6 denotes a difference of two between the class specific item difficulty parameters of items 2 to 6, and EVEN denotes a difference of plus or minus one between the class specific item difficulties of the even items. The parameter sets are chosen to create a condition where the sumscore distribution in each latent class is more or less similar (EVEN), and a condition where the sumscore distribution differs between the latent classes (2TO6).

Item difficulty parameters: set 2TO6										
Latent Class 1	0.0	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
Latent Class 2	0.0	0.0	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0
Item difficulty parameter: set EVEN										
Latent Class 1	0.0	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
Latent Class 2	0.0	-1.0	-1.5	-2.0	-0.5	1.0	0.5	0.0	1.5	3.0

Table 1: Simulation parameters

For every parameter set × number of subjects combination (2TO6-400, 2TO6-2000, EVEN-400, and EVEN-2000) one hundred data sets are simulated. Subsequently six collateral variables are appended to each data set. The collateral variables differed in the strength of association with the latent class variable. Table 2 lists the six levels of association between the latent class and the collateral variables in terms of bivariate probabilities. Note that the column 'names' (50/50,60/40,$\cdots$) correspond to the conditional probabilities, these will be used as shorthand for the collateral variables ('equal' denotes equality between the latent class and collateral variable).

Association Latent class × Collateral variable												
	50/50		60/40		70/30		80/20		90/10		equal	
LC 1	.25	.25	.30	.20	.35	.15	.40	.10	.45	.05	.50	.00
LC 2	.25	.25	.20	.30	.15	.35	.10	.40	.05	.45	.00	.50

Table 2: Proportions in Latent class × collateral variable cross classification

Each data set is subsequently analyzed without a collateral variable using model equation (4), and with each of the six collateral variables separately using model equation (5). Both models can comfortably be specified and analyzed with the LEM program (Vermunt, 1997).

5 Results

First the item parameter estimates are evaluated, by looking at the mean and standard deviation of the deviances, which we define as the difference between the simulation parameters and the estimated parameters. Next, a few statistics concerning the algorithm are reported, and finally the possibility to assign response patterns to the latent classes is evaluated. This is done by assigning response patterns to one of the latent classes using the (log) ratio of the posterior probabilities.

5.1 Parameter estimates

To evaluate the quality of item parameter estimates, differences between the simulation values and the estimated values (deviances) were computed. A difficulty arises because of the possibility of converging to local solutions. Some extreme deviances were obtained. We could, of course, have used the simulation parameters as starting values, but then information about the computational procedure, number of iterations and susceptibility of converging to local maxima, would be lost. As an alternative, we let the algorithm converge to local solutions (using random starting values). To not let the extreme deviances distort the evaluation of the parameter estimates so called trimmed means and standard deviations, with the highest and lowest fifteen percent cut off, will be presented in the following.

First the means of the deviances are presented in Figure 1 for the main and interaction parameters, corresponding with δ_j and Δ_j respectively.

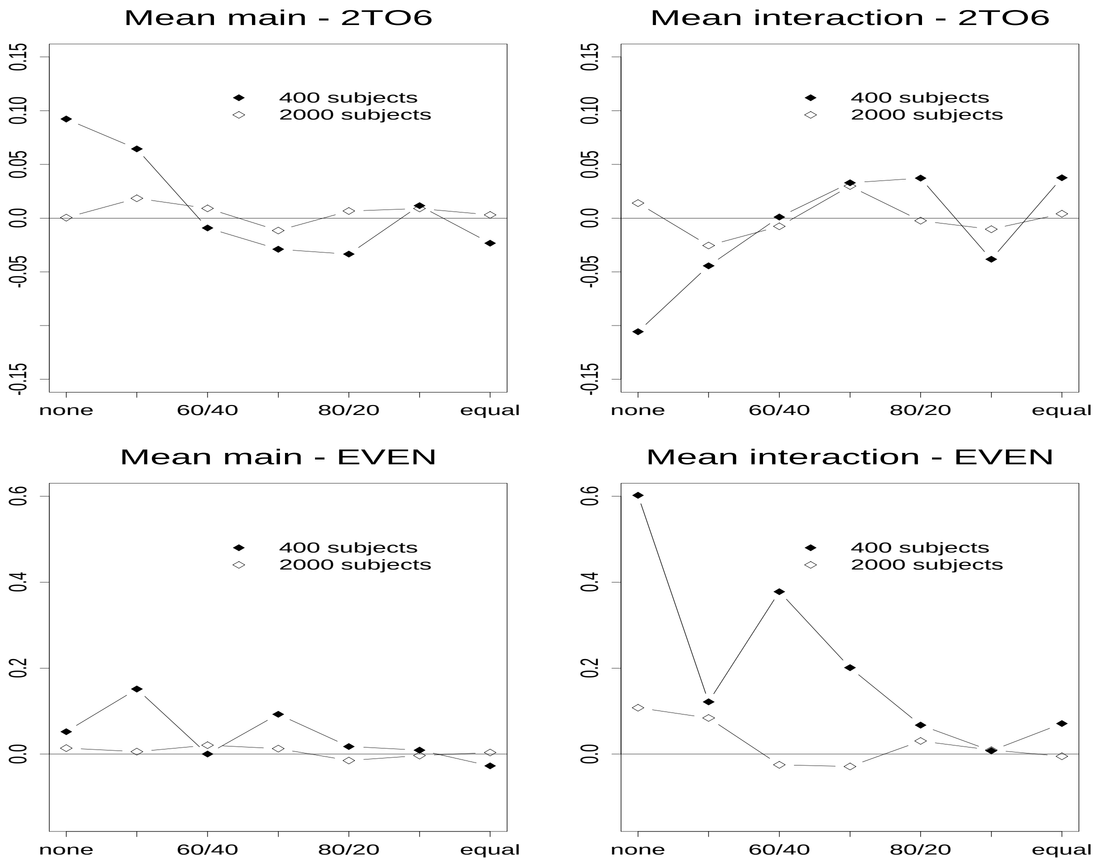

Figure 1: Mean difference of the simulation parameters and estimated parameters. Obtained by aggregating over all items and replications within every –condition × number of subjects × data set × parameter type– combination.

Note that the y-axes are not on the same scale. The figure roughly suggests two trends. The parameters are better estimated when there are more subjects, which is to be expected.

The bandwidth seems to become smaller when stronger collateral variables are incorporated in the model, especially when the sample is smaller. This finding was expected because the variance of the parameter estimates was expected to reduce (the variance of the mean equals the variance divided by the number of cases).

Next the standard deviations of the deviances are presented in Figure 2 for the main and interaction parameters, corresponding with δ_j and Δ_j respectively.

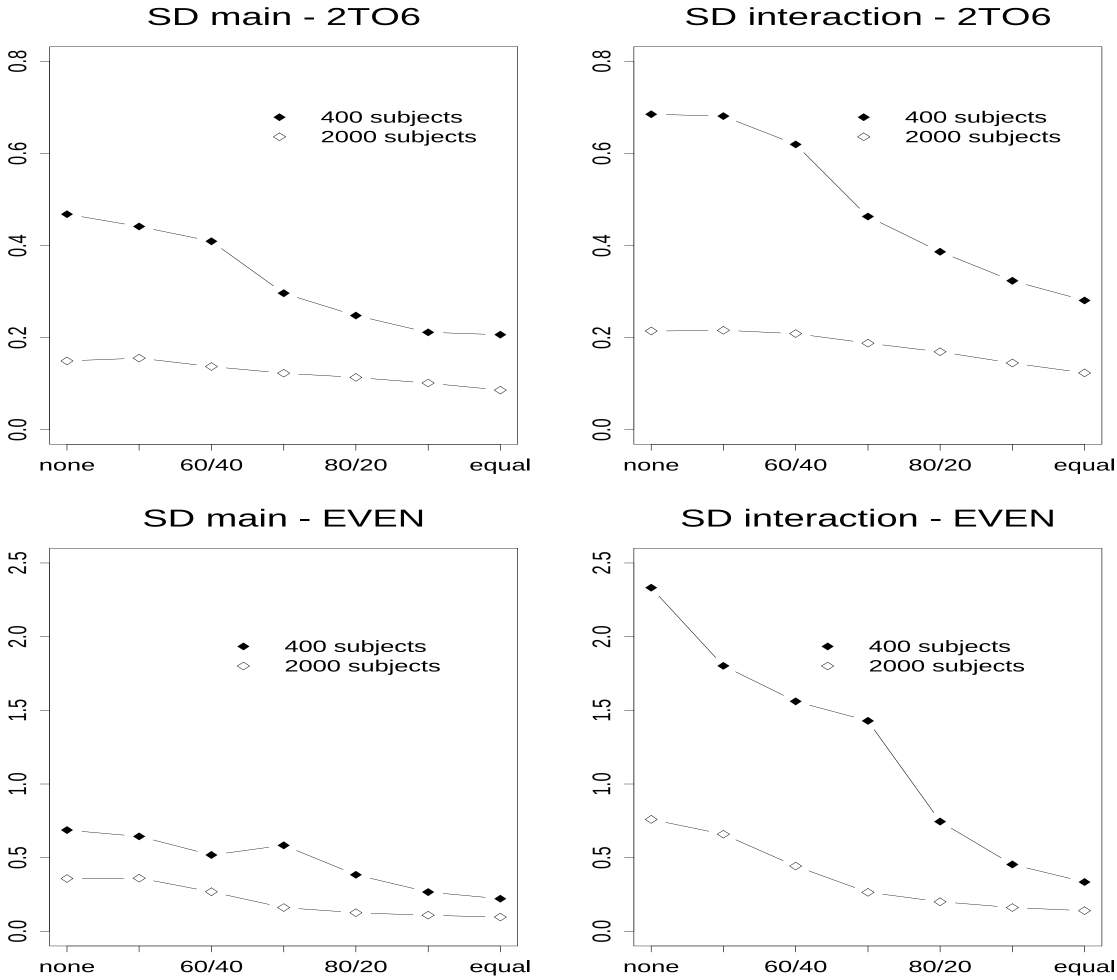

Figure 2: Standard deviation of the difference of the simulation parameters and estimated parameters. Obtained by aggregating over all items and replications within every –condition × number of subjects × data set × parameter type– combination.

Note again the difference in scale of the y-axes. The figure shows the expected reduction in variance when more informative collateral variables are used, or when more subjects are used. Furthermore, the interaction parameters are less accurately estimated than the main parameters, and the parameters of data sets in the EVEN condition are less accurately estimated than the parameters of data sets in the 2TO6 condition. This last finding is due to the 'separability' of the latent classes for the different data sets. Recall that data sets generated with parameter set 2TO6, as opposed to data sets generated with parameter set EVEN, have a different sumscore distribution for the latent classes. Stated differently, members of the two latent classes are further apart (on the sumscore marginal) in the observed contingency table.

5.2 Convergence

In Figure 3 the 25^{th}, 50^{th}, and 75^{th} percentile of the number of iterations needed to converge are reported.

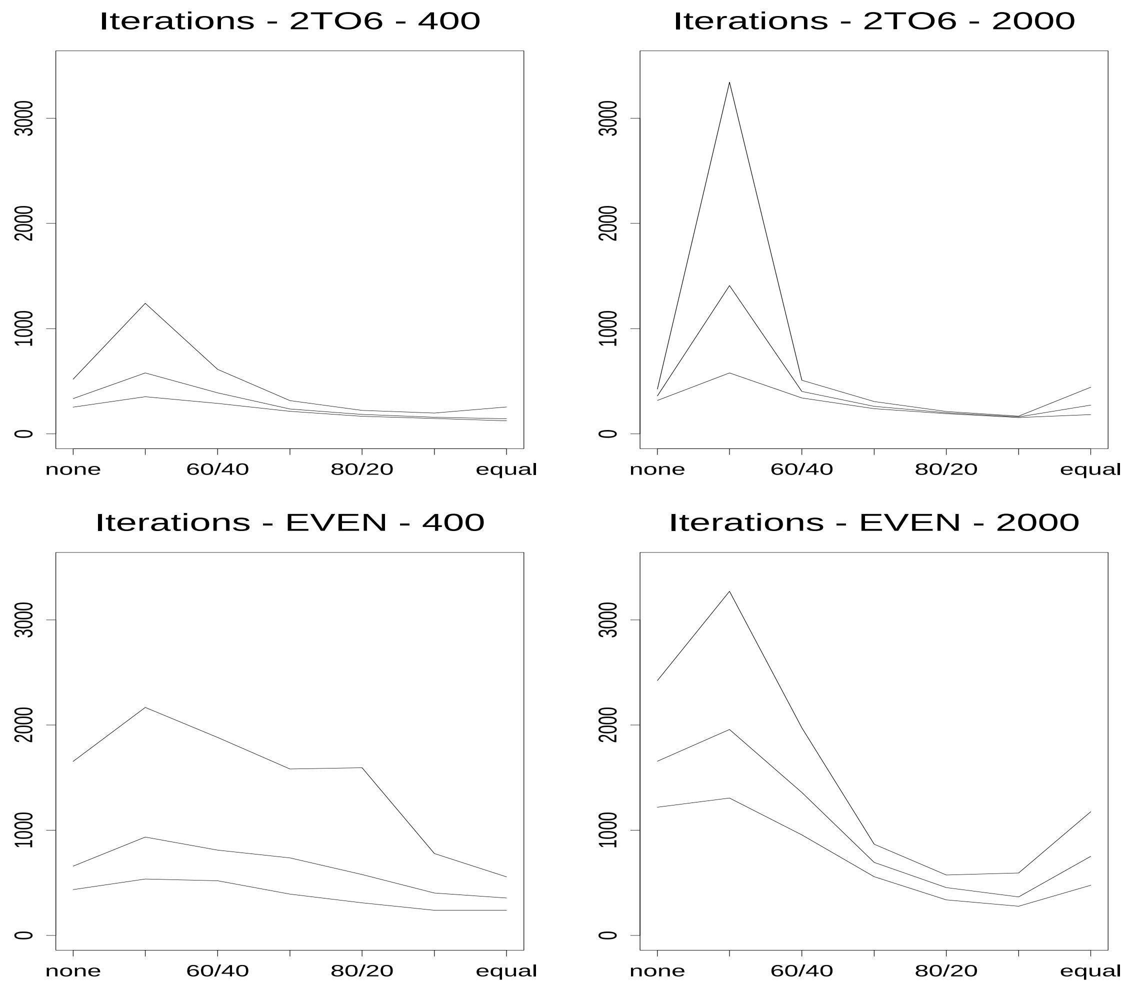

Figure 3: The 25^{th} - 50^{th} - 75^{th} percentiles of the number of iterations. Obtained by aggregating over all replications within every –condition × number of subjects × data set–combination.

Note that, in the condition where the collateral variable coïncides with the latent class variable, the conditional probabilities are (ideally) estimated as either 1 or 0 and the corresponding interaction parameter would tend to infinity or minus infinity. This condition thus deviates from the general trend that the number of iterations decreases as the association between the latent class variable and the collateral variable becomes stronger. The increase for the condition in which the latent class and the collateral variable are independent is also not surprising. Since the collateral variable contains no information about the latent class, the interaction between the collateral variable and the latent class should tend to zero. This corresponds to the situation where a collateral variable cannot influence the parameters because it is absent. So an analysis without a collateral variable can be thought of as an analysis with a collateral variable in which the interaction parameter between the latent class and the collateral variable is implicitly fixed to zero, the correct value.

The cost of incorporating a collateral variable is that the number of computations per iteration increases and that more memory is needed to store the necessary values. Fortunately, the number of computations per iteration and the storage requirements increase only slightly, because the collateral variable can only influence the Mixed Rasch model via the latent class marginal. In general, it took longer to converge if no collateral variable was used, except when the latent class and collateral variable were independent.

The expectation that the collateral variable renders the algorithm less susceptible for converging to 'improper' solutions is evaluated by simply counting the number of extreme deviances. An absolute deviance of two or more is counted as extreme. Counts that are much bigger than expected, indicate convergence to a local solutions. The expected counts can be computed by integrating the normal distribution with the standard deviations from Figure 2 over the interval $[-2, 2]$. Using this criterion all but a few counts have expected value greater than zero. The counts are reported in Table 3 together with their rounded expected value if greater than zero.

	2TO6				EVEN			
	Main		Interaction		Main		Interaction	
	400	2000	400	2000	400	2000	400	2000
none	50	0	88	0	98	39	227 (352)	94 (8)
50/50	45	0	86	0	71	44	191 (240)	85 (2)
60/40	43	0	72	0	46	26	174 (180)	48
70/30	8	0	17	0	66	2	143 (145)	1
80/20	1	0	1	0	32	0	88 (7)	0
90/10	0	0	1	0	13	0	31	0
equal	0	0	0	0	2	0	4	0

Table 3: The number of absolute deviances greater that two. The rounded expected counts (given the empirical trimmed standard deviations) are in braces when greater then 0.

From this table we see that the number of extreme parameter values exceeds what can be expected, for the conditions with weak association between the latent class and collateral variable, indicating convergence to local extremes. The high expected counts for the interaction parameters in data sets simulated with parameter set EVEN indicates that the problem with local maxima is such that a thirty percent cutoff still leaves a few outliers (causing large standard deviations and thus large expected counts).

5.3 Class membership assignment

Finally the ability to assign patterns to latent class membership is evaluated. For every response pattern posterior probabilities can be computed for the latent classes. We used the ratio of these probabilities to assign *all* subjects to one of the latent classes. Note that the ratio might be very close to one. In practical situations, when the quality of the classification is critical, an interval might be constructed for this ratio around one, subjects are not assigned to either of the classes when the ratio falls within this interval (to close to one). In Figure 4 the percentage rightly classified subjects is reported. The classification is done using the item pattern only, and using both the item pattern and the collateral variable.

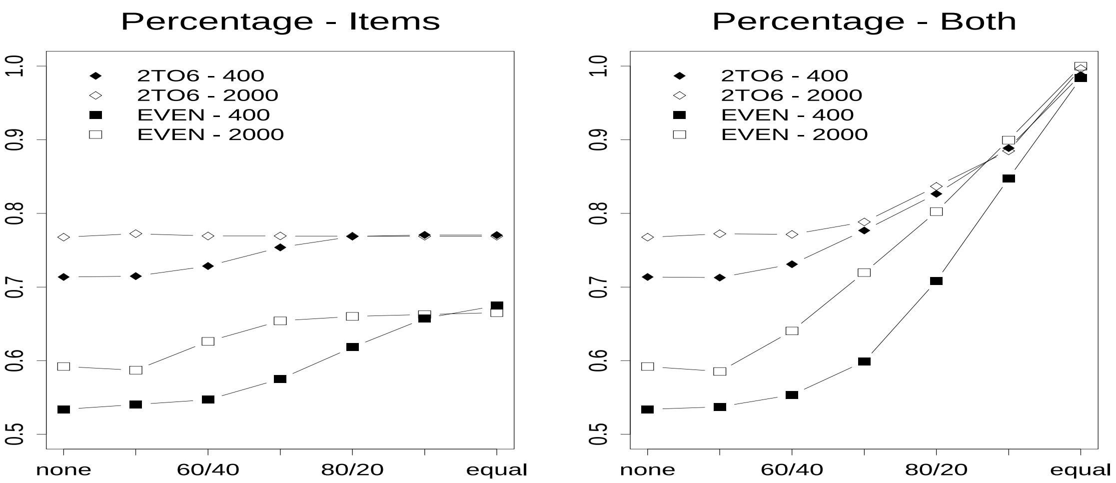

Figure 4: Percentage of correct class assignments, using the item pattern only and using both the item pattern and the collateral variable. Aggregated over all replications within every –condition × number of subjects– combination.

The percentage rightly classified is bounded when using the items only. This is because an item pattern is assigned to one latent class, while it is perfectly feasible for two subjects from different latent classes to have the same item pattern. The left figure shows that the bound is approached if the number of subjects increases, or if the association between the latent class and the collateral variable becomes stronger (due to better parameter estimates and less improper solution). The right figure shows that as the collateral information becomes better the posterior probabilities are dominated by this variable, and thus approach high percentages of rightly classified subjects.

6 Conclusion

In social science, uni-dimensionality of measurement instruments is hard (if not impossible) to achieve. In practical psychometric research it is standard procedure to test for bias effects of sexe, race etc. These observed bias variables can be no more than indicators of the true underlying variable that causes the differential item functioning. The theoretical elegance of a Rasch Model within latent classes is hard to deny. A weakness of this model is that identification can only be achieved when the probability structures are sufficiently different between latent classes, otherwise sample sizes must become unreasonably large. Here the incorporation of collateral variables in the Mixed Rasch Model comes into play. External variables that contain information about the latent class could be incorporated into the model. For the data sets generated in this simulation the gain of one strongly associated collateral variable (in terms of standard errors) can be the same, or even exceed, a fivefold increase in sample size.

As a matter of fact, nothing prevents us from using more than one collateral variable. The cost in terms of computational procedure is rather small. It might be possible that a few weakly associated collateral variables can jointly contain enough information to reduce the variance of the parameter estimates considerably.

References

Böckenholt, U. (1997). Concomitant variables in latent change models. In R. L. Jürgen Rost (Ed.), *Applications of latent trait and latent class models in the social sciences* (pp. 361–369). Waxmann Verlag GmbH.

Clogg, C. C., & Goodman, L. A. (1984). Latent structure analyses of a set of multidimentional contingency tables. *Journal of the American Statistical Association*(79), 762–771.

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

Fisher, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, *36*, 359–374.

Fisher, G. H., & Molenaar, I. W. (1995). *Rasch models - foundations, recent developments, and applications.* New York: Springer.

Gupta, S., & Chintagunta, P. K. (1994). On using demographic variables to determine segment membership in logit mixture models. *Journal of Marketing Research*(31), 128–136.

Heinen, T. (1993). *Discrete latent variable models.* Tilburg: University Press.

Kelderman, H., & Macready, G. B. (1990). The use of loglinear models for assessing differential item functioning across manifest and latent examinee groups. *Journal of Educational Measurement*, *27*(4), 307–327.

Linden, W. Van der, & Hambleton, R. (1997). *Handbook of modern item response theory.* Berlin: Springer.

Little, R. A., & Rubin, D. (1987). *Statistical analyses with missing data.* John Wiley & Sons.

Mislevy, R. J., & Sheehan, K. M. (1989a). Information matrices in latent-variable models. *Journal of Educational Statistics*, *14*(4), 335–350.

Mislevy, R. J., & Sheehan, K. M. (1989b). The role of collateral information about examinees in item parameter estimation. *Psychometrika*, *54*(4), 661–679.

Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, *55*(2), 195–215.

Orchard, T., & Woodbury, M. A. (1972). A missing information principle: Theory and applications. In L. M. Le Cam, J. Neyman, & E. L. Scott (Eds.), *Proceedings of the sixth berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 697–715). University of California Press.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, *14*, 271-282.

Rost, J. (1991). A logistic mixture distribution model for polychotomous item responses. *British Journal of Mathematical & Statistical Psychology*, *44*, 75-92.

Rost, J., & Langeheine, R. (1997). *Applications of latent trait and latent class models in the social siences.* Waxmann Verlag GmbH.

Vermunt, J. (1997). *Log-linear models for event histories.* Thousand Oaks, CA, USA: Sage Publications, Inc.

Wilson, M., Draney, K., & Engelhard, J., G. (1999). *Objective measurement: Theory into practice* (Vol. 5). Norwood, NJ: Ablex.

Wilson, M., Engelhard, J., G., & Draney, K. (1997). *Objective measurement: Theory into practice* (Vol. 4). Norwood, NJ: Ablex.