



Compensatory and non-compensatory multidimensional randomized item response models

Jean-Paul Fox¹, Rinke Klein Entink² and Marianna Avetisyan^{1*}

¹University of Twente, The Netherlands

²TNO Zeist, The Netherlands

Randomized response (RR) models are often used for analysing univariate randomized response data and measuring population prevalence of sensitive behaviours. There is much empirical support for the belief that RR methods improve the cooperation of the respondents. Recently, RR models have been extended to measure individual unidimensional behaviour. An extension of this modelling framework is proposed to measure compensatory or non-compensatory multiple sensitive factors underlying the randomized item response process. A confirmatory multidimensional randomized item response theory model (MRIRT) is proposed for the analysis of multivariate RR data by modelling the response process and specifying structural relationships between sensitive behaviours and background information. A Markov chain Monte Carlo algorithm is developed to estimate simultaneously the parameters of the MRIRT model. The model extension enables the computation of individual true item response probabilities, estimates of individuals' sensitive behaviour on different domains, and their relationships with background variables. An MRIRT analysis is presented of data from a college alcohol problem scale, measuring alcohol-related socio-emotional and community problems, and alcohol expectancy questionnaire, measuring alcohol-related sexual enhancement expectancies. Students were interviewed via direct or RR questioning. Scores of alcohol-related problems and expectancies are significantly higher for the group of students questioned using the RR technique. Alcohol-related problems and sexual enhancement expectancies are positively moderately correlated and vary differently across gender and universities.

1. Introduction

In sample surveys, it can be difficult to obtain reliable information on stigmatizing or socially undesirable/unacceptable matters using the common direct questioning

*Correspondence should be addressed to Marianna Avetisyan, Department of Research Methodology, Measurement and Data Analysis, University of Twente, Faculty of Behavioural Sciences, Drienerlolaan 5, 7522 NB Enschede, Enschede, Overijssel, P.O. Box 217, 7500 AE Enschede, The Netherlands (e-mail:m.avetisyan@gw.utwente.nl).

procedure. Direct questioning on sensitive matters often leads to refusals, non-responses, or socially desirable answers. Warner (1965) developed the randomized response (RR) technique to gather information on such sensitive matters by protecting the privacy of the respondents. It is shown in several studies (e.g., Lensvelt-Mulders, Hox, van der Heijden, & Maas, 2005) that the cooperation of respondents improved due to the RR technique. But despite the evident usefulness of the RR technique, inferences from applications utilizing it are limited to estimating population proportions. Further, the traditional RR models (e.g., Greenberg, Abul-Ela, Simmons, & Horvitz, 1969; Warner, 1965) are only appropriate for the analysis of univariate RR data, they do not account for individual response probabilities, and they do not allow for heterogeneity across respondents. In many cases, outcome data are multivariate or correlated, and it is appealing to model the individual outcomes while taking account of the dependency structure.

To motivate the problem, questionnaire data to measure different psycho-social dimensions of problem drinking among college students were collected on 793 students from four colleges/universities in North Carolina. Furthermore, responses to alcohol expectancy questionnaire items were collected, which measure alcohol-related sexual enhancement expectancies. A part of the participants were questioned via an RR technique to investigate whether this increases the accuracy of self-reports of sensitive information on the different latent dimensions. Further, interest is focused on relationships between latent factors, and their relationship with background variables (e.g., age, gender, racial origin).

Several attempts have been made to extend the class of RR models by modelling the item response process and/or by including various sources of information such as ancillary variables. Scheers and Dayton (1988) model the relation between the population proportion with the sensitive characteristic and covariate information. They showed that the use of relevant covariate information improved the estimation of the population proportion with the sensitive characteristic. Böckenholt and van der Heijden (2007) and Böckenholt, Barlas, and van der Heijden (2009) proposed an item randomized response model for multivariate dichotomous RR data that accounts for the possibility (1) that not all respondents may follow the randomization instructions (say 'no' regardless of the question), and (2) that respondents provide intentionally misleading answers to conceal socially undesirable behaviour. Fox (2005) and Fox and Wyrick (2008) modelled multivariate dichotomous and polytomous data with a randomized item response theory (IRT) model that accounts for individual differences in the response process, and that enables the computation of individual response probabilities, and the measurement of a unidimensional underlying sensitive characteristic. De Jong, Pieters, and Fox (2010) developed a mixture randomized IRT model for polytomous randomized responses.

The IRT modelling approach to RR data is generalized to measure multiple latent sensitive characteristics together with relationships with background variables using a structural multivariate regression model, given dichotomous or polytomous randomized responses. The situation is considered in which multiple latent factors underlie the manifest randomized responses in a compensatory or non-compensatory way. This study extends the work of Böckenholt and van der Heijden (2007), who proposed a between-item multidimensional item randomized response model for binary response data. In their study, multiple item bundles are considered, where each item bundle is used to measure a specific construct using the Rasch model in a non-compensatory way. At the level of observations, responses are assumed to be conditionally independently distributed given one of the factors. The present generalization makes it possible to handle also polytomous randomized item responses, compensatory items when the response process involves multiple constructs, and more advanced item response models that allow item-specific

discriminations/loadings for multiple factors. A full Bayesian estimation procedure is proposed which supports the joint estimation of all parameters, including the estimation of all factor loadings. Here, the mixture model component for dealing with non-compliant response behaviour will not be considered.

In the spirit of multidimensional confirmatory item-factor models, a Bayesian multidimensional confirmatory IRT model is proposed for dichotomous and polytomous data to measure and relate factors underlying the individual sensitive characteristics given randomized response observations. The unobservable factors can be interpreted as a combination of subscale components, or as compensatory or non-compensatory factors that influence the item probabilities in a combined way. This modelling approach connects with recent developments in multidimensional IRT research showing the computational feasibility and increasing attention in the methodology (e.g., Chambers, 2010; Edwards, 2010; Reckase, 2009; Wirth and Edwards, 2007).

The proposed model consists of three components. At the first stage the multivariate RR data are related to individuals' response probabilities via an RR model. At the second stage, the response process is modelled by assuming a multidimensional IRT model for the underlying true responses, which would have been observed if the responses had not been randomized. This enables the measurement of individual response probabilities and individual latent sensitive characteristics. At the third stage, the latent sensitive characteristics are considered to be outcomes of a multivariate regression model. This enables a marginal interpretation for the individual outcomes while appropriately accounting for the dependency structure. The multivariate model has the advantage that the dependency structure can be described parsimoniously in terms of correlation coefficients of the underlying latent characteristics. That is, heterogeneity across respondents and across groups can be properly modelled.

A Markov chain Monte Carlo (MCMC) algorithm is developed to estimate all parameters simultaneously. It is shown that the posterior computation can proceed through a Gibbs sampling algorithm using auxiliary variables. Two augmentation steps facilitate a sampling-based approach for estimating all model parameters simultaneously. First, discrete variables are defined that represent the true item responses that would have been observed without randomizing responses. The conditional distribution of the latent true item responses given the randomized responses are derived via Bayes' theorem. Second, normally distributed latent variables are defined that are manifested as discrete true item responses through a threshold specification. The developed algorithm generalizes the procedure of Fox and Wyrick (2008) and De Jong *et al.* (2010) to deal with the multidimensional IRT model and the structural multivariate latent variable component.

In the following the three-stage multidimensional randomized item response theory model (MRIRT) model is presented. Then a general MCMC algorithm is described for dichotomous and polytomous RR data. Different prior choices are discussed that lead to proper posterior distributions. Then the posterior computations are illustrated with a simulation study. Subsequently, a description is given of the joint Bayesian multidimensional IRT analysis of the College Alcohol Problem Scale and Alcohol Expectancy Questionnaire RR data with careful attention to the underlying factor structure. Finally, the pros and cons of the new model are discussed and briefly compared with other approaches in the literature.

2. Modelling individual response probabilities

In general, the RR technique is used to estimate the proportion, π , of respondents belonging to a sensitive class in the population. Horvitz, Shah, and Simmons (1967)

proposed an unrelated question RR design that is based on two questions. One provocative question is related to the sensitive characteristic, and the other unrelated question refers to a non-sensitive innocuous attribute. Each respondent selects randomly, by means of a randomizing device such as a die or spinner, one of the two questions and answers it truthfully. The respondent does not tell the interviewer which question he has selected to answer. If the population proportion of the non-sensitive characteristic is known, and this is built into the randomizing device (see Boruch, 1971), the probability of a positive response will be

$$n_1/n = p_1\pi + (1 - p_1)p_2, \quad (1)$$

where n_1 (n_1/n) is the number (proportion) of ‘yes’ answers reported by n individuals, and p_1 is the probability that the randomizing device selects the question related to the sensitive characteristic. Subsequently, with probability $1 - p_1$, the unrelated question is selected and with probability p_2 a positive response is given. Note that parameters p_1 and p_2 are known since they are specified by the randomizing device. Using equation (1), the proportion of persons affirming or denying the intrusive item can be accurately estimated. De Schrijver (2012) compared the forced response technique with the unrelated questioning technique and concluded that the forced response technique was better understood.

For multivariate RR data, let the random variable \mathbf{Y}_i denote the observed randomized responses and the random variable \mathbf{U}_i the true responses that would have been observed if these responses had not been randomized. The distribution of observed randomized responses relates to the distribution of true item responses according to the RR sampling design; that is,

$$P(\mathbf{Y}_i) = P(Y_{i1}, \dots, Y_{iK}) = \prod_k [p_1 P(U_{ik}) + (1 - p_1)p_2]. \quad (2)$$

The true individual response probabilities $P(U_{ik})$ will be modelled, which can improve the estimate of the proportion with the sensitive characteristic in the population depending on the available explanatory information. It will provide information at the individual level since estimates of individual response probabilities can be obtained as well as relationships with explanatory variables.

3. The model

3.1. Probit response functions

It will be assumed that a set of K items are composed to measure a multidimensional latent trait, θ_i , where $\theta_i = (\theta_{i1}, \dots, \theta_{iQ})^t$. Let, θ_q , $q = 1, \dots, Q$, contain the traits of respondents $i = 1, \dots, N$ on dimension q . The true categorical outcome, U_{ik} , represents the item response of person i on item k ($k = 1, \dots, K$). These item responses may be dichotomous or polytomous. For dichotomous item responses a two-parameter Q -dimensional IRT model is used to specify the relation between the level of a latent trait and the probability of a particular item response; that is,

$$P(U_{ik} = 1 \mid \theta_i, \mathbf{A}_k, b_k) = \Phi(\mathbf{A}_k^t \theta_i - b_k) = \Phi\left(\sum_q A_{kq} \theta_{iq} - b_k\right), \quad (3)$$

where \mathbf{A}_k is the vector of item discrimination parameters or factor loadings, and b_k is the item difficulty parameter. The item parameters will also be denoted by ξ_k , with

$\xi_k = (\mathbf{A}_k, b_k)$. For polytomous item responses, the probability that an individual obtains a grade c ($c = 1, \dots, C$) on item k is defined by a graded response model

$$\begin{aligned} P(U_{ik} = c | \theta_i, \mathbf{A}_k, \boldsymbol{\kappa}_k) &= \Phi(\mathbf{A}_k^t \theta_i - \kappa_{kc-1}) - \Phi(\mathbf{A}_k^t \theta_i - \kappa_{kc}) \\ &= \Phi\left(\sum_q A_{kq} \theta_{iq} - \kappa_{kc-1}\right) - \Phi\left(\sum_q A_{kq} \theta_{iq} - \kappa_{kc}\right), \end{aligned} \quad (4)$$

where the boundaries between the response categories are represented by an ordered vector of thresholds $\boldsymbol{\kappa}$; let $\xi_k = (\mathbf{A}_k, \boldsymbol{\kappa}_k)$. There are a total of $C - 1$ threshold parameters and Q discrimination parameters for each item. For the logistic IRT model replace $\Phi(\cdot)$ with $\mathcal{L}(\cdot)$. Although the polytomous IRT model in equation (4) also comprehends the two-parameter IRT model, the two-parameter IRT model is presented separately according to equation (3). Reckase (2009) gives an overview of multidimensional IRT models.

3.2. Forced randomized response design

According to the forced RR design, two probabilities are specified by the randomizing device: the probability p_1 that the respondent has to answer the sensitive question, and the probability p_2 that a forced positive response is given. Using equations (2) and (3), the probability of a positive randomized response equals

$$\begin{aligned} P(Y_{ik} = 1 | \theta_i, \mathbf{A}_k, b_k) &= p_1 P(U_{ik} = 1 | \theta_i, \mathbf{A}_k, b_k) + (1 - p_1) p_2 \\ &= p_1 \Phi(\mathbf{A}_k^t \theta_i - b_k) + (1 - p_1) p_2. \end{aligned} \quad (5)$$

This is easily extended to multiple (say, $c = 1, \dots, C$) response categories. The randomizing device determines if the item is to be answered honestly with probability p_1 or a forced response is scored in category c with probability $(1 - p_1) p_2(c)$. Using equations (2) and (4), the probability of scoring in category c equals

$$\begin{aligned} P(Y_{ik} = c | \theta_i, \mathbf{A}_k, \boldsymbol{\kappa}_k) &= p_1 P(U_{ik} = c | \theta_i, \mathbf{A}_k, \boldsymbol{\kappa}_k) + (1 - p_1) p_2(c) \\ &= p_1 [\Phi(\mathbf{A}_k^t \theta_i - \kappa_{kc-1}) - \Phi(\mathbf{A}_k^t \theta_i - \kappa_{kc})] + (1 - p_1) p_2(c). \end{aligned} \quad (6)$$

3.3. Structural multivariate latent model

In multidimensional IRT, it is often assumed that the latent traits θ_i , $i = 1, \dots, N$, are exchangeable and multivariate normally distributed. That is, $\theta_i \sim N_Q(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$, where the population means of latent traits and variations and dependencies among latent trait dimensions do not depend on any individual information. This model specification can be extended to include clustering of background information of respondents. In general, a multivariate model makes it possible to investigate correlations between latent traits, its dependence on the individual and the group level, to test simultaneously effects of an explanatory variable on several latent traits, and to test for differential effects of explanatory variables on various latent traits.

Let S predictor variables for respondent i be stored in $\mathbf{X}_i^t = (X_{i1}, \dots, X_{iS})$. A structural multivariate latent model for the latent traits is expressed by

$$\begin{pmatrix} \theta_{11} & \cdots & \theta_{1Q} \\ \theta_{21} & \cdots & \theta_{2Q} \\ \vdots & \ddots & \vdots \\ \theta_{N1} & \cdots & \theta_{NQ} \end{pmatrix} = \begin{pmatrix} X_{11} & \cdots & X_{1S} \\ X_{21} & \cdots & X_{2S} \\ \vdots & \ddots & \vdots \\ X_{N1} & \cdots & X_{NS} \end{pmatrix} \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1Q} \\ \gamma_{21} & \cdots & \gamma_{2Q} \\ \vdots & \ddots & \vdots \\ \gamma_{S1} & \cdots & \gamma_{SQ} \end{pmatrix} + \begin{pmatrix} e_{11} & \cdots & e_{1Q} \\ e_{21} & \cdots & e_{2Q} \\ \vdots & \ddots & \vdots \\ e_{N1} & \cdots & e_{NQ} \end{pmatrix},$$

that is,

$$\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{e}, \quad (7)$$

where \mathbf{e} is an $N \times Q$ matrix whose rows are independently distributed, $\mathbf{e}_i \sim N_Q(\mathbf{0}, \boldsymbol{\Sigma}_\theta)$. The $S \times Q$ matrix $\boldsymbol{\gamma}$ contains unknown fixed effects parameters. The latent traits on the Q dimensions for each respondent have a covariance matrix $\boldsymbol{\Sigma}_\theta$ but the latent traits are uncorrelated across individuals. A complete overview of multivariate multiple regression models can be found in Johnson and Wichern (2002).

The within-subject covariance matrix $\boldsymbol{\Sigma}_\theta$ can be modelled as functions of unknown covariance parameters, where each structure for $\boldsymbol{\Sigma}_\theta$ can have important subject-specific interpretations. The efficiency of the regression parameter estimates may be improved by modelling the covariance matrix parsimoniously as independent, with a constant variance parameter across subjects' latent trait dimensions,

$$\boldsymbol{\Sigma}_\theta = \sigma^2 \mathbf{I}_Q, \quad (8)$$

where \mathbf{I}_Q is a $Q \times Q$ identity matrix.

Other cases arise by allowing the variance σ^2 to vary across latent trait dimensions, or by allowing the variance to vary across groups of subjects. In both cases $\boldsymbol{\Sigma}_\theta$ remains a diagonal matrix. The general extension is the unstructured covariance matrix with $Q(Q+1)/2$ parameters, and possibly with covariance parameters that vary across groups.

Finally, in a multivariate mixed effects modelling approach, variation can be modelled in latent trait dimensions within and between groups of subjects, for example, due to treatment effects. Following Schafer and Yucel (2002), let n_j ($j = 1, \dots, J$) denote the number of subjects in cluster j , and let \mathbf{W}_j be an $n_j \times R$ design matrix linking to the random effect parameters $\boldsymbol{\zeta}_j$, an $R \times Q$ matrix of coefficients specific to subjects in cluster j . With these random effects, equation (7) is generalized to

$$\boldsymbol{\theta}_j = \mathbf{X}_j \boldsymbol{\gamma} + \mathbf{W}_j \boldsymbol{\zeta}_j + \mathbf{e}_j, \quad (9)$$

where \mathbf{X}_j is an $n_j \times S$ matrix of explanatory variables. The n_j rows of \mathbf{e}_j are assumed to be independently distributed as $N_Q(\mathbf{0}, \boldsymbol{\Sigma}_\theta)$. The random effects are independently distributed as $\text{vec}(\boldsymbol{\zeta}_j) \sim N(\mathbf{0}, \mathbf{T})$, where the $\text{vec}(\cdot)$ operator creates a column vector from the columns of $\boldsymbol{\zeta}_j$. It is assumed that for clusters j and j' , $\text{cov}(\mathbf{e}_j, \mathbf{e}_{j'}) = 0$, $\text{cov}(\boldsymbol{\zeta}_j, \boldsymbol{\zeta}_{j'}) = 0$, and $\text{cov}(\mathbf{e}_j, \boldsymbol{\zeta}_{j'}) = 0$. Then the covariance matrix for $\text{vec}(\boldsymbol{\theta}_j)$ equals

$$(\mathbf{I}_Q \otimes \mathbf{W}_j) \mathbf{T} (\mathbf{I}_Q \otimes \mathbf{W}_j)^t + \boldsymbol{\Sigma}_\theta \otimes \mathbf{I}_{n_j}. \quad (10)$$

A well-known structure is the case where each \mathbf{W}_j is a vector of ones that gives the between-within mixed model structure called 'compound symmetry'. If modelling a large

number of latent traits it may be advantageous to restrict \mathbf{T} to a block-diagonal structure, indicating that there are no a priori associations between the columns of ζ_j .

3.4. Identification issues

The MRIRT model can be identified by imposing restrictions on some of the parameters. Following Béguin and Glas (2001), one approach is to fix the mean and covariance matrix of θ to zero and the identity matrix, respectively. To avoid the rotational invariance, constraints are necessary for some elements of \mathbf{A} , for example, $A_{kq} = 0$ for $k = 1, \dots, Q - 1$ and $q = 2, \dots, Q$. Furthermore, at least one element in each column of \mathbf{A} is constrained to be positive (see also Lopes & West, 2004).

It is also possible to consider the covariance parameters as unknowns by imposing the restrictions $A_{kq} = 1$ if $k = q$, and $A_{kq} = 0$ if $k \neq q$ for $k = 1, \dots, Q$ and $q = 1, \dots, Q$, and setting the mean equal to zero. Instead of restricting the upper-diagonal elements of the matrix of factor loadings, the diagonal elements of Σ_θ can also be restricted to one such that only the non-diagonal elements are free parameters.

4. Bayesian inference

An MCMC procedure is proposed for model estimation. This MCMC algorithm is based on developed MCMC methods for multidimensional IRT and factor-analytic models (e.g., Béguin & Glas, 2001; Bolt & Lall, 2003; Jackman, 2001; Lopes & West, 2004; Sheng & Wikle, 2007; Shi & Lee, 1998; Song & Lee, 2001; Yao & Boughton, 2007). A straightforward MCMC implementation is not possible, since discrete randomized response data are observed. Therefore, following the MCMC method of Fox and Wyrick (2008) for unidimensional IRT, a double augmentation step is proposed to sample the discrete true item response data and continuous latent response data given the RR data. As a result, the joint distribution of the parameters and augmented data are considered to circumvent a direct evaluation of the likelihood function, which is computationally intensive.

The joint posterior distribution can be expressed as

$$\begin{aligned} p(\mathbf{u}, \theta, \zeta, \mathbf{A}, \kappa, \gamma, \Sigma_\theta, \mathbf{T} | \mathbf{y}) &\propto p(\mathbf{y} | \mathbf{u}) p(\mathbf{u} | \theta, \mathbf{A}, \kappa) p(\theta | \gamma, \Sigma_\theta, \zeta) p(\zeta | \mathbf{T}) \\ &\quad p(\mathbf{A}) p(\kappa) p(\mathbf{T}) p(\Sigma_\theta) p(\gamma) \\ &\propto \prod_j \left[\prod_i \left[\prod_k p(y_{ijk} | u_{ijk}) p(u_{ijk} | \theta_i, \mathbf{A}_k, \kappa_k) \right] \right. \\ &\quad \left. p(\theta_i | \gamma, \zeta_j, \Sigma_\theta) \right] p(\zeta_j | \mathbf{T}) p(\mathbf{A}) p(\kappa) p(\mathbf{T}) p(\Sigma_\theta) p(\gamma), \end{aligned}$$

where $p(y_{ijk} | u_{ijk})$ defines the RR process given the characteristics of the randomizing device. The term $p(u_{ijk} | \theta_i, \mathbf{A}_k, \kappa_k)$ defines the multidimensional IRT component for the true response data, and $p(\theta_i | \gamma, \zeta_j, \Sigma_\theta)$ the multivariate model component for the factors.

The structural multivariate parameters and the multidimensional IRT parameters are sampled from the full conditionals, where only the threshold parameters κ are sampled using a Metropolis–Hastings step. The full conditionals are easily derived given the augmented data. Therefore, the randomized response data \mathbf{Y} will be augmented with latent data \mathbf{U} and \mathbf{Z} . The random variable U_{ijk} represents the true response to item k of a person indexed ij . The conditional probability of a true response U_{ijk} given a randomized

response Y_{ijk} is considered. Let $\pi_{ijk}(c)$ denote the probability of responding in category c according to the IRT model in equation (5) or (6); it follows that

$$\begin{aligned} P(U_{ijk} = c | Y_{ijk} = c; \pi_{ijk}(c)) &= \frac{P(U_{ijk} = c, Y_{ijk} = c; \pi_{ijk}(c))}{P(Y_{ijk} = c; \pi_{ijk}(c))} \\ &= \frac{\pi_{ijk}(c)[p_1 + (1 - p_1)p_2]}{\pi_{ijk}(c)p_1 + (1 - p_1)p_2}, \end{aligned}$$

where p_1 and p_2 are the known randomizing device characteristics. In general, the conditional distribution of a true item response given a randomized item response is a multinomial with cell probabilities

$$\Delta(c) = \frac{\pi_{ijk}(c)p_1 I(c = c') + \pi_{ijk}(c)(1 - p_1)p_2}{\pi_{ijk}(c')p_1 + (1 - p_1)p_2}, \quad (11)$$

for $c, c' = 1, \dots, C_k$. For binary response data, the conditional distribution is Bernoulli with the success probability defined in (11) and $c = 1$.

Subsequently, the full conditional of the augmented data \mathbf{Z} is a normal distribution,

$$\mathbf{Z}_{ijk} | U_{ijk}, \boldsymbol{\theta}_i, \boldsymbol{\xi}_k \sim \begin{cases} N(\mathbf{A}_k^t \boldsymbol{\theta}_i - b_k, 1) & \text{for binary data,} \\ N(\mathbf{A}_k^t \boldsymbol{\theta}_i, 1) & \text{for polytomous data,} \end{cases} \quad (12)$$

with U_{ijk} the indicator of Z_{ijk} being positive for binary response data, and $U_{ijk} = c$ if $\kappa_{kc-1} \leq Z_{ijk} \leq \kappa_{kc}$ for polytomous response data.

The prior for the non-fixed factor loadings is standard normal. The full conditional for \mathbf{A}_k is normal with mean $(\boldsymbol{\theta}^t \boldsymbol{\theta} + \mathbf{I}_Q)^{-1}(\boldsymbol{\theta}^t(\mathbf{Z}_k + b_k))$ and variance $(\boldsymbol{\theta}^t \boldsymbol{\theta} + \mathbf{I}_Q)^{-1}$. When \mathbf{A}_k contains fixed elements, the vector of free random loadings is sampled given the other vector of fixed loadings. The difficulty parameters are not included in the mean term when dealing with polytomous data, but thresholds are introduced as additional restrictions on the space spanned by the augmented data vectors.

Non-informative proper priors for the difficulty or threshold parameters are specified. The threshold parameters in equation (4) are assumed to be independent and uniformly distributed subject to the condition $\kappa_{k0} < \kappa_{k1} < \dots < \kappa_{kC_k}$, with $\kappa_{k0} = -\infty$ and $\kappa_{kC_k} = \infty$. Sampled values from the conditional distribution of the threshold parameters can be obtained using the Metropolis–Hastings algorithm; see Fox (2005, 2010) for specific details. For binary data, the full conditional posterior of the difficulty parameter is normal with mean

$$\frac{-\sum_i (Z_{ik} - \mathbf{A}_k^t \boldsymbol{\theta}_i) / N + \mu_b / \sigma_b}{N^{-1} + \sigma_b^{-1}} \quad (13)$$

and variance $(N^{-1} + \sigma_b^{-1})^{-1}$ using a normal prior with mean μ_b and variance σ_b . The mean and variance parameters are sampled from a normal and an inverse-gamma distribution, respectively, given the difficulty parameters.

Consider the multivariate regression model in equation (7) for the latent variables $\boldsymbol{\theta}$ and the multidimensional IRT model in equation (3). Subsequently, the full conditional distribution is normal with mean

$$\theta_{ij} | \mathbf{Z}_{ij}, \mathbf{A}, \mathbf{b}, \gamma, \Sigma_\theta \sim N((\mathbf{A}^t \mathbf{A} + \Sigma_\theta^{-1})^{-1} \hat{\theta}_{ij}, (\mathbf{A}^t \mathbf{A} + \Sigma_\theta^{-1})^{-1}), \quad (14)$$

where $\hat{\theta}_{ij} = (\mathbf{A}^t (\mathbf{Z}_{ij} + \mathbf{b}) + \Sigma_\theta^{-1} \gamma^t \mathbf{X}_{ij})$.

4.1. Structural multivariate parameters

The posterior distributions of the fixed effects parameters γ in equation (7) follow from standard Bayesian linear model results. The prior distribution for the multivariate regression parameter is assumed to be normal with fixed mean and variance parameter γ_0 and Γ , respectively. It follows that

$$\gamma | \theta, \gamma_0, \Sigma_\theta, \Gamma \sim N(\tilde{\gamma}, \Sigma_\gamma),$$

with

$$\begin{aligned} \tilde{\gamma} &= \Sigma_\gamma ((\Sigma_\theta^{-1} \otimes \mathbf{X}^t \mathbf{X}) \hat{\gamma} + \Gamma^{-1} \gamma_0) \\ \Sigma_\gamma &= ((\Sigma_\theta^{-1} \otimes \mathbf{X}^t \mathbf{X}) + \Gamma^{-1})^{-1}, \end{aligned}$$

Subsequently, the covariance matrix Σ_θ has an inverse-Wishart distribution with $N + n_0$ degrees of freedom and scale matrix \mathbf{S}_θ^{-1} , with $\mathbf{S}_\theta = \sum_j (\theta_j - \mathbf{X}_j \gamma) (\theta_j - \mathbf{X}_j \gamma)^t + \mathbf{I}_Q$ and $n_0 \geq Q$, to specify a non-informative proper prior distribution. Note that all parameters of the covariance matrix Σ_θ are free parameters when using identifying restrictions on the item parameters and factor loadings.

The posterior distribution of the fixed effects parameters and the within-subject covariance matrix given a multivariate mixed effects model according to equation (9) are slightly different due to the fact that the error terms are defined as $\mathbf{e}_j = \theta_j - \mathbf{X}_j \gamma - \mathbf{W}_j \zeta_j$, and by allowing variation across clusters. According to (9), it follows that the posterior distribution of the random effect parameters is a product of normal distributions. As a result,

$$\text{vec}(\zeta_j) | \theta_j, \Sigma_\theta, \gamma, \mathbf{T} \sim N(\text{vec}(\hat{\zeta}_j), \Omega_\zeta), \quad (15)$$

where

$$\begin{aligned} \text{vec}(\hat{\zeta}_j) &= \Omega_\zeta (\Sigma_\theta^{-1} \otimes \mathbf{W}_j^t) \text{vec}(\theta_j - \mathbf{X}_j \gamma) \\ \Omega_\zeta &= (\mathbf{T}^{-1} + \Sigma_\theta^{-1} \otimes \mathbf{W}_j^t \mathbf{W}_j)^{-1}. \end{aligned}$$

Subsequently, when \mathbf{T} is block diagonal, independent draws from a inverse-Wishart distribution with degrees of freedom $\nu + J$ and scale parameter $\Lambda_\zeta + \sum_j \zeta_{jq} \zeta_{jq}^t$, where ζ_{jq} denotes the q th column of ζ_j . The hyperparameter ν is usually set to R to specify a disperse prior, and a priori independence is assumed among the random effects, that is, $\Lambda_\zeta = \mathbf{I}_R$.

4.2. Implementation issues

The MCMC algorithm can handle RR data as well as direct questioning data since the properties of the randomizing device are known and corresponding parameters can be set to specific values. For direct questioning data, p_1 is set to one. This corresponds with the

approach of Chaudhuri and Mukerjee (1988) who permitted an option for direct questioning for those who volunteer to reveal the truth viewing the attribute not stigmatizing enough.

Ignorable missing response values are handled by sampling latent augmented data Z_{ijk} without truncating the values to a specific domain but based on given values of the item parameters and the latent variable. This imputation-based procedure creates a complete data set and the procedure is easily implemented in the MCMC algorithm. The imputed augmented values have larger standard deviations since they are not restricted to a specific domain such that uncertainty due to missing values is taken into account.

The convergence of the MCMC algorithm is depending on several factors. Convergence can be slow when the amount of missing information is high. In that case the latent person parameters and item parameters are poorly estimated with large variances. Also more iterations might be needed in order to obtain stable parameter estimates. Convergence can also be slow when the number of clusters is large, because for large J the posterior distribution for \mathbf{T} given ζ becomes very tight, and, as a result, a drawn value of the covariance matrix \mathbf{T} may remain close to its previous value. Convergence can be informally assessed by examining trace plots, time series plots, plots of the average of each parameter across multiple chains, and plots of the running average. Formal and informal convergence diagnostics can be found in Brooks and Gelman (1998) and Gilks (1996). Starting values for the MCMC algorithm can be obtained by fitting a multidimensional IRT model to the data but ignoring the randomized response character of the discrete response data using the MCMC algorithm of Béguin and Glas (2001). However, it will be shown in a simulation study that convergence properties of the proposed MCMC algorithm are good and independent of chosen starting values.

5. Simulation study

In this section, results are reported from a simulation study for parameter recovery based on the MRIRT model for randomized item response data.

Data were simulated for two correlated factors ($Q = 2$), where a total of $N = 750$ subjects responded to $K = 20$ items. For the two factors a structural multivariate model was assumed with two explanatory variables. The covariance between factors was set at 0.30, and the factor variance components were both 1. The explanatory variable values were generated from a standard normal distribution. Four response categories were assumed, with the threshold parameters for all items set at $\kappa_C = (-1, 0, 1)$. For the forced RR part, the probabilities were set at $p_1 = .667$ and $p_2 = .25$. To avoid rotation of the two factors, the first item only loaded on the first factor, and the second item only on the second factor. Furthermore, for reasons of identification the mean and variance of each factor were restricted to the simulated values.

Using randomly generated starting values, two MCMC chains of 10,000 iterations each were simulated, and the first 5,000 iterations were discarded for the burn-in. Using the CODA package in R, convergence of the algorithm was assessed using several statistics as well as the visual inspection of trace plots. Figure 1 shows trace plots of two chains for the regression coefficients. For the MCMC chains of the regression parameters, Gelman's R statistic was estimated to be 1.01, which is below the recommended threshold of 1.10. Also for other parameters those statistics suggested convergence of the algorithm. The partial autocorrelation function for several chains suggested a first-order Markov process, only for some chains showing a statistically significant, but minor, autocorrelation of around .20 at lag 2.

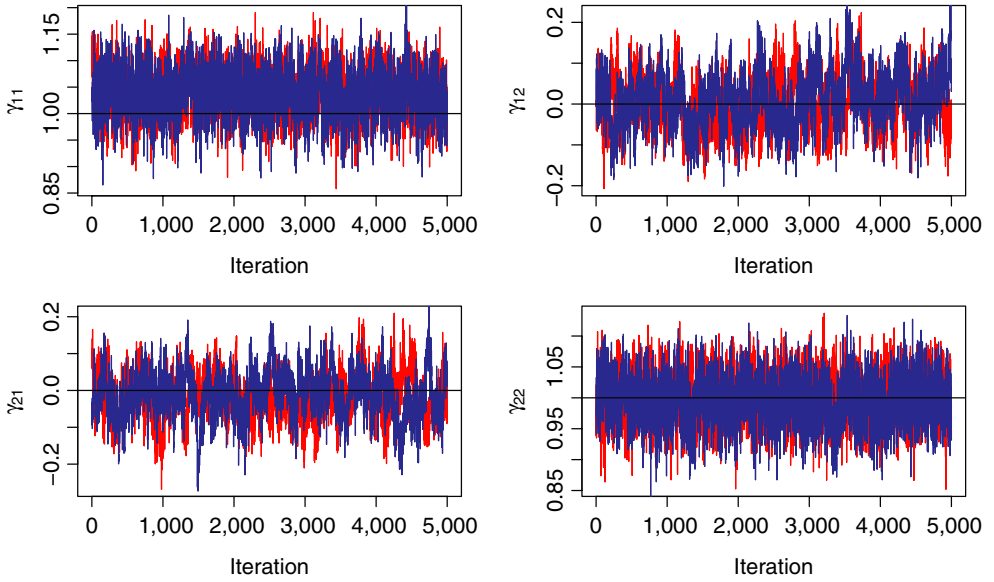


Figure 1. Trace plots of estimated regression coefficients showing two MCMC runs, where the solid black line represents the simulated true value.

Table 1. Results of simulation study. Generating values, means and standard errors of recovered values

Fixed	Gen. Coeff.	RIRT model		
		Mean	SD	95% HPD
<i>First factor</i>				
γ_{10}	0	–	–	–
γ_{11}	1.00	1.04	0.04	[0.95,1.12]
γ_{12}	0.00	0.00	0.07	[–0.13,0.13]
<i>Second factor</i>				
γ_{20}	0	–	–	–
γ_{21}	0.00	–0.02	0.06	[–0.15,0.10]
γ_{22}	1.00	1.00	0.04	[0.92,1.08]
<i>Random</i>				
	Coeff.	Mean	SD	95% HPD
$\Sigma_{\theta_{11}}$	1.00	0.92	0.09	[0.79,1.06]
$\Sigma_{\theta_{12}}$	0.30	0.35	0.10	[0.17,0.51]
$\Sigma_{\theta_{22}}$	1.00	1.01	0.08	[0.88,1.14]

Recovery of the loadings and the factor values was assessed graphically (not shown) and showed that the true against the re-estimated parameters closely followed the identity line. Finally, Table 1 shows the true and re-estimated parameters for the structural multivariate model on the subject parameters, indicating that the MCMC algorithm worked well for estimation of the model.

6. Measuring drinking problems and alcohol-related expectancies among college students

Thirteen items of the College Alcohol Problem Scale (CAPS; O'Hare, 1997) and four items of the Alcohol Expectancy Questionnaire (AEQ; Brown, Christiansen, & Goldman, 1987) were used to assess alcohol problems and alcohol-related expectancies among college students. The questionnaire items are given in the Appendix. The sensitive nature of the study supports an RR questioning technique to avoid refusals and misleading responses concealing socially undesirable behaviour. Any self-reported information about negative consequences of drinking is likely to be biased due to socially desirable responding. It is investigated whether the RR technique improved the cooperation of the respondents and the accuracy of the self-report data by comparing the RR outcomes with the direct questioning outcomes. Furthermore, using the MRIRT modelling approach, multiple sensitive constructs underlying both scales (CAPS and AEQ) are measured and their relationships with background information analysed.

The AEQ is used to measure the degree of expectancies associated with drinking alcohol. Expectancies related to alcohol use are known to influence alcohol use and behaviour while drinking (e.g., Werner, Walker, & Greene, 1995). The entire test consists of 90 items and covers six dimensions (see Brown *et al.*, 1987). In the present study, attention was focused on alcohol use-related sexual enhancement expectancies using four items covering sexual enhancement expectancies, which are given in the Appendix.

The CAPS instrument is one of the major self-report measures used to assess drinking problems. The items cover socio-emotional (e.g., hangovers, memory loss, depression) as well as community problems (e.g., driving under the influence, engaging in activities related to illegal drugs, problems with the law). O'Hare (1997) developed the CAPS instrument to measure different psycho-social dimensions of problem drinking among college students. The two factors, socio-emotional and community problems, were identified from a factor analysis which explained more than 60% of the total variance. Fox and Wyrick (2008) analysed the CAPS data using a unidimensional RIRT model to measure a general construct alcohol dependence. Although the model described the data well, a multidimensional approach provides insight in the different factors related to problem drinking, factor-specific relationships with background variables, and supports the multidimensional nature of the CAPS.

Data were collected through a survey study in 2002 at four local colleges/universities, Elon University ($N = 495$), Guilford Technical Community College ($N = 66$), University of North Carolina ($N = 166$), and Wake Forest University ($N = 66$). A total of 351 students was assigned to the direct questioning (DQ) condition and 442 to the RR condition. Students in the DQ group served as the study's control group and were instructed to answer the questionnaire as they normally would. Students in the RR condition used a spinner to assist them in completing the questionnaire. For each item of the CAPS and AEQ, the spinner was used as a randomizing device and the outcome determined whether to answer honestly or to register a forced response. The properties of the spinner were set such that an honest answer was requested with a probability of 60% and a forced response with a probability of 40%. When a forced response was to be generated, each response was given an equal probability of 20%. No identifying information was collected, but age, gender, and ethnicity were also registered. Each class of students (5–10 participants) was randomly assigned to the DQ group or the RR group and selected from the same population. It was not possible to randomly assign students.

The following MRIRT model was used to analyse the response data:

$$\begin{aligned}
 P(Y_{ik} = c | \boldsymbol{\theta}_i, \mathbf{A}_k, \boldsymbol{\kappa}_k) &= p_1 \pi_{ik} + (1 - p_1) p_2(c), \\
 \pi_{ik} &= \Phi(\mathbf{A}_k^t \boldsymbol{\theta}_i - \kappa_{i,(c-1)}) - \Phi(\mathbf{A}_k^t \boldsymbol{\theta}_i - \kappa_{i,c}), \\
 \boldsymbol{\theta}_i &= \boldsymbol{\gamma}_0 + \boldsymbol{\gamma}_1 RR_i + \mathbf{e}_i,
 \end{aligned}$$

where $\mathbf{e}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\theta)$ and RR_i equals one for each dimension q when student i was assigned to the RR group and zero otherwise. According to the forced RR sampling design, $p_1 = .60$ and $p_2(c) = .20$, for $c = 1, \dots, 5$; in the DQ design $p_1 = 1$. The factor loadings \mathbf{A} and item thresholds are assumed to be independent of the questioning technique. The MRIRT model was identified by fixing the mean in each dimension, such that $\boldsymbol{\gamma}_0 = \mathbf{0}$. The variance was identified by restricting the variance components of each factor to 1. The so-called rotational variance was identified by assigning Q items uniquely to Q dimensions. Each model was estimated using 50,000 MCMC iterations using a burn-in period of 10,000 iterations for parameter estimation. The convergence of chains was inspected using plots and various MCMC convergence diagnostics.

6.1. Multidimensional scale analysis

First, a two-factor RIRT model was estimated, where items 1 and 14 had just one free non-zero loading to identify two factors. In Figure 2, the estimated factor loadings are given for the two-factor RIRT model stated in equation (16). The estimated factor loadings are standardized by dividing each loading by the average item loading. For each component, the sign of the loadings is set in such a way that a higher latent score corresponds to a

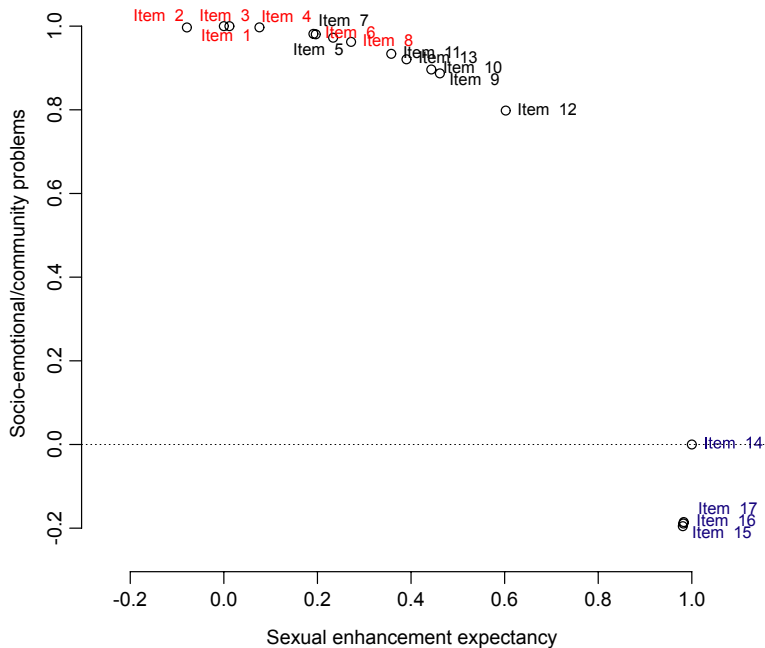


Figure 2. Estimated standardized factor loadings for the two-factor RIRT model

higher observed score. It can be seen that items 14–17 measure the factor alcohol-related expectancy and that most other items are clearly measuring another factor, socio-emotional/community problems, which was labelled alcohol dependence in the Fox and Wyrick (2008) analysis. The loadings are all above .75, which indicates that both factors can be interpreted. Note that expectancies are increasing with alcohol consumption and slightly diminish socio-emotional/community problems given the negative factor loadings for the other component.

Second, a three-factor RIRT model was estimated to investigate whether problems associated with drinking were represented by two factors (i.e., socio-emotional and community problems), and sexual enhancement expectancies by another factor. Besides items 1 and 14, the loadings of item 5 were also restricted to identify the community problems factor as reported in the literature. In Table 2, the standardized estimated factor loadings of the three factors are given. Items 1–4, 6, 8, and 9 associate with the first factor, representing socio-emotional problems, and have factor loadings higher than .60. This first component represents drinking-related problems including depression, anxiety, and trouble with family, where the problems will increase with alcohol consumption. Some of the items also associate with the two other components. The second component, labelled community problems, covers items 5, 7, and 10–13, with loadings higher than .60, except for item 12. As reported in the literature, this item is associated with the community problems factor but relates also to the other components and most strongly with the third component. The community problems factor covers acute physiological effects of drunkenness together with illegal and potentially dangerous activities (e.g., driving under the influence).

For the two-factor RIRT model, the second and third response options of the AEQ items were more likely to be endorsed than the second and third response options of the CAPS

Table 2. CAPS-EA scale: estimated weighted factor loadings for the three-factor analysis

Subscale Items	Three-factor RIRT model		
Socio-emotional	Comp. 1	Comp. 2	Comp. 3
1 Feeling sad, blue or depressed	1.00	.00	.00
2 Nervousness or irritability	.99	.12	-.05
3 Hurt another person emotionally	.94	.33	.08
4 Family problems related to your drinking	.82	.55	.15
6 Badly affected friendship or relationship	.83	.49	.26
8 Caused other to criticize your behavior	.78	.50	.38
9 Nausea or vomiting	.73	.39	.58
Community problems	Comp. 1	Comp. 2	Comp. 3
5 Spent too much money on drugs	.00	1.00	.00
7 Hurt another person physically	.49	.83	.24
10 Drove under the influence	.44	.73	.52
11 Spent too much money on alcohol	.61	.65	.46
12 Feeling tired or hung over	.59	.41	.69
13 Illegal activities associated with drug use	.08	.95	.30
Alcohol expectancy scale	Comp. 1	Comp. 2	Comp. 3
14 I often feel sexier after I've had a couple of drinks	.00	.00	1.00
15 I'm a better lover after a few drinks	-.11	-.06	.99
16 I enjoy having sex more if I've had some alcohol	-.16	-.01	.99
17 After a few drinks, I am more sexually responsive	-.16	-.02	.99

items. The corresponding threshold estimates of the AEQ items in comparison to the CAPS items for response categories 2 and 3 are lower – except for CAPS item 12 (feeling tired or hung over) on which students scored relatively high. The AEQ items 14–17 and CAPS item 12 can be considered as the less severe items. Items with relatively high thresholds, item 4 (family problems related to drinking), item 5 (spending too much money on drugs), item 7 (hurting another person physically), and 13 (illegal activities associated with drug use) were severe, where most students responded almost never or seldom. The threshold estimates of the three-factor model were similar, except the threshold estimates of item 5 were not very stable and much higher. This follows from the fact that most responses above one were considered to be forced randomized responses. Item 5 with a very low prevalence did not provide much information to assess drinking problems. A prior restriction on the upper bound led to a numerically stable solution.

6.2. Structural model analysis

In the two- and three-factor RIRT models, the multivariate latent factor model was extended with an explanatory variable called RR and an indicator variable called Female (which equals 1 when true). For each dimension, both explanatory variables were included.

In Table 3, the structural multivariate parameter estimates of the three-factor and two-factor RIRT models are given. For the two-factor model, loadings of items 1 and 14 were fixed to identify two factors. One overall factor represents a composite measure of alcohol-related problems (i.e., socio-emotional and community problems) and the other factor alcohol-related sexual enhancement expectancies. It can be seen that there is a moderate positive covariance of 0.65 between the two factors, where the component variances are slightly smaller than 1.

Table 3. CAPS-EA scale: Parameter estimates of two- and three-factor RIRT model

Parameter	Two-factor		Three-factor	
	Mean	SD	Mean	SD
Socio-emotional/community				
γ_{11} (RR)	.20	.09	.21	.10
γ_{21} (Female)	.01	.06	.05	.07
Alcohol expectancy				
γ_{12} (RR)	.22	.06	.21	.07
γ_{22} (Female)	.03	.04	.06	.05
Community				
γ_{13} (RR)			.32	.10
γ_{23} (Female)			-.30	.09
Variance parameters	Mean	SD	Mean	SD
$\Sigma_{\theta_{11}}$	0.96	0.05	0.98	0.05
$\Sigma_{\theta_{12}}$	0.65	0.07	0.55	0.06
$\Sigma_{\theta_{13}}$			0.35	0.08
$\Sigma_{\theta_{22}}$	0.98	0.05	1.06	0.05
$\Sigma_{\theta_{23}}$			0.37	0.08
$\Sigma_{\theta_{33}}$			0.99	0.07
Information criteria				
–2log-likelihood		20,757		19,678

The students in the RR condition score significantly higher for both factors. For the RR group, the average latent scores are .20 and .22 on the composite drinking problem scale and the alcohol-related expectancy scale, respectively, which are both zero in the DQ group. Fox and Wyrick (2008), who performed a unidimensional RIRT analysis using only the CAPS items, reported an RR effect of .23. The present multidimensional approach shows a comparable RR effect, as it does for the AEQ scores. The females and males show comparable scores on both factors.

In the three-factor RIRT model, problems associated with drinking are represented by two factors (i.e., socio-emotional and community problems) and sexual enhancement expectancies by another factor. The RR effects are significantly different from zero for all three factors, where the effect on the factor representing community problems related to alcohol use is around .32 and slightly higher than the effects on the other factors, which are around .21. It seems that students were less willing to admit to alcohol-related community problems, which induced more socially desirable responses than the other

Table 4. CAPS-EA scale: Differences across colleges and universities using the two- and three-factor RIRT model

Parameter	Two-factor		Three-factor	
	Mean	SD	Mean	SD
Socio-emotional/community				
γ_{11} (RR)	.29	.08	.28	.09
<i>School variables</i>				
γ_{21} (Elon)	.19	.06	.19	.07
γ_{31} (UNCG)	-.19	.10	-.14	.11
γ_{41} (Wake Forest)	-.23	.12	-.11	.14
γ_{51} (Guilford)	.24	.12	.05	.13
Alcohol expectancy				
γ_{12} (RR)	.19	.06	.19	.07
<i>School variables</i>				
γ_{22} (Elon)	.04	.05	.04	.05
γ_{32} (UNCG)	-.11	.07	-.10	.08
γ_{42} (Wake Forest)	-.23	.08	-.23	.09
γ_{52} (Guilford)	.30	.09	.29	.11
Community				
γ_{13} (RR)			.30	.09
<i>School variables</i>				
γ_{23} (Elon)			.12	.06
γ_{33} (UNCG)			-.13	.10
γ_{43} (Wake Forest)			-.37	.13
γ_{53} (Guilford)			.38	.13
Variance parameters				
	Mean	SD	Mean	SD
$\Sigma_{\theta_{11}}$	0.94	0.05	0.96	0.05
$\Sigma_{\theta_{12}}$	0.62	0.07	0.55	0.06
$\Sigma_{\theta_{13}}$			0.45	0.07
$\Sigma_{\theta_{22}}$	0.97	0.05	0.99	0.05
$\Sigma_{\theta_{23}}$			0.47	0.06
$\Sigma_{\theta_{33}}$			0.94	0.05
Information criteria				
-2log-likelihood		20,858		19,660

factors. The relatively high thresholds of items 5, 7, and 13 measuring community problems indicated that they were more severe and most likely more sensitive. Male students scored significantly higher than the female students on the factor representing community problems related to alcohol use. That is, male students were more likely to experience alcohol-related community problems than females. This gender effect was not found for the other factors.

Interest was focused on the average student drinking problems and expectancies of the four selected colleges/universities that took part in the experiment. The clustering of students in colleges/universities was represented using effect coding. In Table 4 the parameter estimates are given for the two-factor and three-factor models. For each factor, the intercept represents the average score across colleges and universities, which was set to zero. In the two-factor model, the average score of the RR group is .29 and .19, both significantly higher than zero. It follows that the mean scores of the alcohol-related problems factor for Guilford Technical Community College and Elon University are significantly higher than the means for University of North Carolina and Wake Forest. For the alcohol-related expectancy factor, Guilford Technical Community College scored on average higher than the other colleges and universities.

For the three-factor model, the mean scores of the three factors of the RR group are significantly higher than zero and comparable when controlling for differences across universities and colleges. It follows that Guilford Technical Community College has the highest average score of alcohol-related community problems and of alcohol-related sexual enhancement expectancies. The results show that alcohol-related sexual enhancement expectancies and community problems are positively correlated, where scores differ across universities and colleges. The estimates of the RR effect indicate that the RR group scored significantly higher in comparison to the DQ group on each subscale. Although validation data are not available, it is to be expected that the RR technique led to an improved willingness of students in answering truthfully, given the random assignment to DQ and RR conditions.

7. Discussion

In educational and psychological measurement, it is often more realistic to assume that multiple constructs influence the performance on test items. The multidimensional item response theory model can be used to assess the underlying latent variable structure given the test results (Reckase, 2009). When surveying sensitive topics, direct questioning may lead to social desirability bias. Therefore, in combination with a randomized response design, a multidimensional randomized item response model is proposed to measure multiple sensitive constructs given multivariate randomized response data. The confirmatory multidimensional model presented can handle dichotomous and polytomous randomized item responses. The application shows a model belonging to the class of compensatory models. However, when every item measures one construct a non-compensatory MRIRT model can be stated in a similar way.

Markov chain Monte Carlo methods have developed to tackle the high-dimensional integration problem in confirmatory multidimensional IRT analysis (e.g., Edwards, 2010; Sheng, 2010). Cai (2010) proposed a Metropolis–Hastings Robbins–Monro algorithm for an exploratory analysis given polytomous response data. Béguin and Glas (2001) developed a full Gibbs sampling procedure and proposed several posterior predictive checks. Yao and Boughton (2007) showed MCMC estimation of multidimensional partial

credit models and the assessment of subscale scores. The present MCMC algorithm for the estimation of the MRIRT model is based on a double augmentation scheme to deal with the categorical randomized response outcomes. Furthermore, the MCMC algorithm can also handle the estimation of the multivariate structural model parameters. This includes the structural regression parameters, which specify the relationship between the multiple sensitive constructs and the explanatory background information, and the correlation structure among the latent variables. The MCMC algorithm has been developed in R and will be made freely available via the internet.

The present survey study on alcohol-related sexual enhancement expectancies and drinking problems showed that randomized response questioning improved the cooperation of the respondents and reduced domain-specific social desirability bias. The joint analysis results support the alcohol expectancy theory (e.g., Brown *et al.*, 1987), which states that positive expectancies due to alcohol use lead to more positive initial drinking experiences, leading in turn to more positive expectancies. Here, it was shown that alcohol-related sexual enhancement expectancy scores were positively correlated with subscale scores for alcohol-related socio-emotional and community problems. In the literature, alcohol-related expectancies have been found to be useful in predicting drinking problems and drinking behaviour, and patterns of problematic use (e.g., Werner *et al.*, 1995). The randomized response technique can improve the accuracy of the self-report data and related predictions, while the multidimensional modelling approach can improve the accuracy of subscale scores by using the additional subscale information (e.g., Yao & Boughton, 2007).

The measurement of alcohol expectancies of individuals is important to identify current and predict future problem drinking. The randomized response technique can improve the quality of the diagnostic self-report data, when respondents tend to underreport alcohol consumption and the negative effects of alcohol use due to social desirability or potential legal ramifications. The multifactor modelling approach will support the multifactorial nature of the expectancy questionnaires and individual measurements of expectancy behaviour given randomized response data.

Acknowledgement

The authors thank Cheryl Haworth Wyrick for providing the data from the study on alcohol use and abuse by college students. Funding for the project was provided by the Netherlands Organization for Scientific Research (NWO).

References

- Béguin, A. A., & Glas, C. A. W. (2001). MCMC estimation of multidimensional IRT models. *Psychometrika*, *66*, 541–562. doi:10.1007/BF02296195
- Böckenholt, U., Barlas, S., & van der Heijden, P. G. M. (2009). Do randomized response designs eliminate response bias? An empirical study of non-compliance behavior. *Journal of Applied Econometrics*, *24*, 377–392.
- Böckenholt, U. & van der Heijden, P. G. M. (2007). Item randomized-response models for measuring noncompliance: Risk-return perceptions, social influences, and self-protective responses. *Psychometrika*, *72*, 245–262.
- Bolt, D. M., & Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo. *Applied Psychological Measurement*, *27*, 395–414.

- Boruch, R. F. (1971). Maintaining confidentiality of data in educational research: A systematic analysis. *American Psychologist*, *6*, 413–430.
- Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.
- Brown, S. A., Christiansen, B. A., & Goldman, M. S. (1987). The alcohol expectancy questionnaire: An instrument for the assessment of adolescent and adult alcohol expectancies. *Journal of Studies on Alcohol*, *48*, 483–491.
- Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis–Hastings Robbins–Monro algorithm. *Psychometrika*, *75*, 33–57.
- Chambers, R. P. (2010). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29.
- Chaudhuri, A., & Mukerjee, R. (1988). *Randomized response: Theory and techniques*. New York: Marcel Dekker.
- Chib, S., & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, *85*, 347–361.
- De Schrijver, A. (2012). Sample survey on sensitive topics: Investigating respondents' understanding and trust in alternative versions of the randomized response technique. *Journal of Research Practice*, *8*(1), M1.
- Edwards, M. C. (2010). A Markov chain Monte Carlo approach to confirmatory item factor analysis. *Psychometrika*, *75*, 474–497.
- Fox, J.-P. (2005). Multilevel IRT using dichotomous and polytomous items. *British Journal of Mathematical and Statistical Psychology*, *58*, 145–172.
- Fox, J.-P. (2010). *Bayesian item response modeling: Theory and applications*. New York: Springer.
- Fox, J.-P., & Wyrick, C. (2008). A mixed effects randomized item response model. *Journal of Educational and Behavioral Statistics*, *33*, 389–415.
- Gilks, W. R. (1996) Full conditional distributions. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice* (pp. 75–88). London: Chapman & Hall.
- Greenberg B. G., Abul-El A., Simmons W. R., & Horvitz D. G. (1969). The unrelated question randomized response model: theoretical framework. *Journal of the American Statistical Association*, *64*, 520–539.
- Horvitz, D. G., Shah, B. V., & Simmons, W. R. (1967). The unrelated question randomized response model. *Proceedings of the American Statistical Association: Social Statistics Section 12*, 65–72.
- Jackman, S. (2001). Multidimensional analysis of roll call data via Bayesian simulation: Identification, estimation, inference, and model checking. *Political Analysis*, *9*, 227–241.
- Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis*. Upper Saddle River: NJ Prentice Hall.
- Lensvelt-Mulders, G. J. L. M., Hox, J. J., van der Heijden, P. G. M., & Maas, C. J. M. (2005). Meta-analysis of randomized response research: Thirty-five years of validation. *Sociological Methods and Research*, *33*, 319–348.
- Lopes, H. F., & West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica*, *14*, 41–67.
- Maddock, J. E., Laforge, R. G., Rossi, J. S., & O'Hare, T. (2001). The college alcohol problems scale. *Addictive Behaviors*, *26*, 385–398.
- O'Hare, T. (1997). Measuring problem drinking in first time offenders: Development and validation of the College Alcohol Problem Scale (CAPS). *Journal of Substance Abuse Treatment*, *14*, 383–387.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.
- Schafer, J. L. & Yucel, R. M. (2002). Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics*, *11*, 437–457.
- Scheers, N. J. & Dayton, C. (1988). Covariate randomized response models. *Journal of the American Statistical Association*, *83*, 969–974.
- Sheng, Y. (2010). Bayesian estimation of MIRT models with general and specific latent traits in Matlab. *Journal of Statistical Software*, *34*(3), 1–27.

- Sheng, Y., & Wikle, C. K. (2007). Comparing multiunidimensional and unidimensional item response theory models. *Educational and Psychological Measurement*, 67, 899–919.
- Shi, J.-Q., & Lee, S.-Y. (1998). Bayesian sampling-based approach for factor analysis models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, 51, 233–252.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis*. London: Sage.
- Song, X.-Y., & Lee, S.-Y. (2001). Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations. *British Journal of Mathematical and Statistical Psychology*, 54, 237–263.
- Warner S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60, 63–69.
- Werner, M. J., Walker, M. D., & Greene, J. W. (1995). Relation of alcohol expectancies to changes in problem drinking among college students. *Archives of Pediatrics & Adolescent Medicine*, 149, 733–739.
- Wirth R. J. & Edwards M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, 12, 58–79.
- Yao, L., & Boughton, K. A. (2007). A multidimensional item response modeling approach for improving subscale proficiency estimation and classification. *Applied Psychological Measurement*, 31, 83–105.

Received 16 December 2012; revised version received 11 March 2013

Supporting Information

The following supporting information may be found in the online edition of the article:

Dataset 1: Compensatory and Noncompensatory Multidimensional Randomized Item Response Analysis of the CAPS and EAQ Data

Dataset 2: Explanation R-Code and additional results. Compensatory and non-compensatory Multidimensional Randomized Item Response Models.

Appendix: CPS-EQ Questionnaire

CAPS: Socio-emotional and community problems

How often (almost always (5), often (4), sometimes (3), seldom (2), almost never (1)) have you had any of the following problems over the past years as a result of drinking too much alcohol?

1. Feeling sad, blue or depressed
2. Nervousness or irritability
3. Hurt another person emotionally
4. Family problems related to your drinking
5. Spent too much money on drugs
6. Badly affected friendship or relationship
7. Hurt another person physically
8. Caused other to criticize your behavior
9. Nausea or vomiting
10. Drove under the influence
11. Spent too much money on alcohol

- 12. Feeling tired or hung over
- 13. Illegal activities associated with drug use

AEQ: Sexual enhancement

- 14. I often feel sexier after I've had a couple of drinks
- 15. I'm a better lover after a few drinks
- 16. I enjoy having sex more if I've had some alcohol
- 17. After a few drinks, I am more sexually responsive