LSAC RESEARCH REPORT SERIES

Bayesian Estimation Methods for Multidimensional Models for Discrete and Continuous Responses with a Structure on the Item and Person Parameters

Cees A. W. Glas
Oksana Korobko
University of Twente, Enschede, The Netherlands

Law School Admission Council
Research Report 07-03
March 2007

A publication of the Law School Admission Council

The Law School Admission Council (LSAC) is a nonprofit corporation whose members are more than 200 law schools in the United States, Canada, and Australia. Headquartered in Newtown, PA, USA, the Council was founded in 1947 to facilitate the law school admission process. The Council has grown to provide numerous products and services to law schools and to more than 85,000 law school applicants each year.

All law schools approved by the American Bar Association (ABA) are LSAC members. Canadian law schools recognized by a provincial or territorial law society or government agency are also members. Accredited law schools outside of the United States and Canada are eligible for membership at the discretion of the LSAC Board of Trustees.

Table of Contents

Executive Summary

In previous Law School Admission Council reports, models for the simultaneous analysis of discrete (correct versus incorrect) and continuous item responses were developed. The current extension of that research focuses on the prediction of item and person parameters from covariates.

The method is illustrated with examples of the analysis of the grades for national school-leaving examinations at the end of secondary education in the Netherlands. In this example, as in one of the previous reports, the interaction between the students' patterns and levels of proficiency and the choice of examination subjects is modeled by extending the basic model for the responses with a model for the choice of the examination subjects that were elective for the students. Using this setup, two research questions are investigated: (a) How much of the variance in the ability parameters for the person is attributable to the schools? and (b) How much of this variance is attributable to gender?

With respect to research question (a), the schools contributed to the students' ability in Language and Economy, but not to their ability in Science. With respect to research question (b), the proportions of variance explained by gender were highest for the ability dimension associated with Language and Science. Further, the average score for ability associated with the Language dimension was higher for female students, while their average on the choice dimension was lower.

Abstract

In previous Law School Admission Council reports, methods for estimating and testing item response theory models were outlined for continuous responses developed in a marginal maximum likelihood framework and a Bayesian framework using a Markov chain Monte Carlo computational method. In the present report, it is shown how item and person parameters can be predicted from covariates. It is assumed that the item and person parameters may have a multilevel structure.

As in the two previous reports, the method is illustrated with examples of the analysis of the grades of central examinations in secondary education in the Netherlands. Further, as in previous reports, the interaction between the students' pattern and level of proficiency and the choice of examination subjects is modeled by enhancing the model for the responses with a model for the choice of the examination subjects.

Introduction

The present Law School Admission Council (LSAC) report should be seen as a follow-up to a previous LSAC report (Glas, 2006) in which a comprehensive Bayesian estimation method using a Markov chain Monte Carlo (MCMC) computational method was developed for simultaneously estimating the parameters of item response theory (IRT) models for discrete and continuous observations. In the present report, the method will be generalized to models with multilevel regression models on the item and person parameters.

As in the previous report, the method will be illustrated with an example of analysis of examination grades of central examinations in secondary education in the Netherlands. Also in the present analysis, the IRT model for the grades is enhanced with a model for the choice of the examination subjects. In the previous report, it was pointed out that ignoring the choice process (i.e., treating the choices as fixed variables) may lead to biased inferences.

In the present analyses, two research questions are investigated: (a) How much of the variance in the latent person parameters is attributable to the schools? and (b) How much of this variance is attributable to gender.

The Model

A Model for Continuous Observations

Let students be indexed $n = 1,\dots,N,$ and let items be indexed $k = 1,\dots,K.$ It is assumed that the observation z_{nk} on student n and item k is normally distributed, that is,

$$P(Z_{nk} = z_{nk} \mid \theta_n, \mathbf{a}_k, b_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \cdot \exp\left(-\frac{(z_{nk} - \eta_{nk})^2}{2\sigma_k^2}\right). \tag{1}$$

The expectation of the item response is a linear function of the explanatory latent variables,

$$\eta_{nk} = \sum_{q=1}^{Q} a_{kq}\theta_{nq} - b_k$$
$$= \mathbf{a}_k'\boldsymbol{\theta}_n - b_k, \tag{2}$$

where $\mathbf{a}_k$ is a vector of the parameters ($a_{k1},\ldots,a_{kq},\ldots,a_{kQ}$), which are usually called factor loadings, and b_k is a location parameter. Further, $\boldsymbol{\theta}_n = (\theta_{n1},\ldots,\theta_{nq},\ldots,\theta_{nQ})$ is the Q-dimensional proficiency parameter of student n. We assume that $\sigma_k^2 = 1$, for all k. That is, we assume that all the observed responses have the same scale.

A Model for Discrete Observations

Glas (2006) considers several IRT models for discrete observations, such as the three-parameter normal ogive model (Birnbaum, 1968; Lord, 1980), the graded response model (GRM, Samejima, 1969), and the sequential model (Tutz, 1990). The theory presented here pertains to all of these models. However, for the application presented here we focus on an IRT model with a single peaked response function. The model will not be used to describe an actual response of a student to an item but a missing data indicator. This missing data indicator is defined as

$$d_{nk} = \begin{cases} 1 & \text{if person } n \text{ responds to item } k \\ 0 & \text{if otherwise.} \end{cases} \tag{3}$$

In the application presented below, the items are examination topics, and the missingness occurs because the students are free to choose a number of examination topics. In the model, it is assumed that students will choose topics with a difficulty that matches their overall ability and that they will avoid topics that are either too easy or too difficult. This will be modeled by an IRT model with a single peaked response function given by

$$P(d_{nk} = 1;\eta_{nk}) = \Phi(\eta_{nk1}) - \Phi(\eta_{nk2}), \tag{4}$$

with $\eta_{nkj} = a_k'\boldsymbol{\theta}_n - b_{kj}$ $(j = 1,2)$ and $b_{k1} < b_{k2}$ to guarantee that $P(Y_{nk} = 1;\eta_{nk})$ is positive. Note that

$$P(d_{nk} = 0;\eta_{nk}) = 1 - \Phi(\eta_{nk1}) + \Phi(\eta_{nk2}). \tag{5}$$

Glas (2006) notes that the model given by (4) is related to a GRM (Samejima, 1969). In the GRM, the probability of a response U_{nk} in category j ($j = 1,\ldots,m$) of item k is given by

$$P(U_{nk} = j;\eta_{nkj}) = \begin{cases} 1 - \Phi(\eta_{nk1}) & \text{if } j = 0 \\ \Phi(\eta_{nkj}) - \Phi(\eta_{nk(j+1)}) & \text{if } 0 < j < m \\ \Phi(\eta_{nkm}) & \text{if } j = m, \end{cases} \tag{6}$$

with $\eta_{nkj} = \boldsymbol{\alpha}_k'\boldsymbol{\theta}_n - \beta_{kj}$. To ensure that the probabilities $P(Y_{nkj} = 1;\eta_{nkj})$ are positive, the restriction $\beta_{k(j+1)} > \beta_{kj}$ for $0 < j < m$ is imposed. Note that the model given by (4) and (5) can be viewed as a GRM for responses that assume the values 0, 1, or 2, where latent responses 0 and 2 are collapsed to an observed response equal to zero. This conceptualization will play a role in the estimation procedure for the model.

The reason for explicitly modeling the missing data indicator rather than treating it as fixed is that the observations and the missing data indicators depend on the same or related latent variables. Holman and Glas (2005) argued that in such cases the so-called ignorability assumption is violated (Rubin, 1976), and in such cases inferences ignoring the response model for the missing data indicator can be severely biased (Bradlow & Thomas, 1998; Holman & Glas, 2005). However, this bias can be removed when the process causing the missingness is modeled concurrently with the observed data using an IRT model (Holman & Glas, 2005; Moustaki & Knott, 1999; Moustaki & O'Muircheartaigh, 2000; O'Muircheartaigh & Moustaki, 1999).

Models for Item and Person Parameters

On a second level, it is assumed that all first-level person parameters are independent samples from multivariate normal distributions. Further, at this point we also assume that the students may be nested under higher-level units. For instance, students may be nested in classes. The higher-level units will be indexed $p = 1, \ldots, P.$ Person parameters $\boldsymbol{\theta}_{np}$ are predicted with a linear regression model, where $\mathbf{X}_{\mathcal{P}np}$ are observed covariates, $\boldsymbol{\beta}_{\mathcal{P}p}$ are the regression parameters in unit $p,$ and $\boldsymbol{\Sigma}_{\mathcal{P}}$ is the covariance matrix of the residuals. So the density of $\boldsymbol{\theta}_{np}$ is given by

$$\boldsymbol{\theta}_{np} \sim N(\mathbf{X}_{\mathcal{P}np}\boldsymbol{\beta}_{\mathcal{P}p}, \boldsymbol{\Sigma}_{\mathcal{P}}).$$

The regression parameters themselves may also be random parameters. We impose a regression model

$$\boldsymbol{\beta}_{\mathcal{P}p} \sim N(\mathbf{W}_{\mathcal{P}p}\boldsymbol{\gamma}_{\mathcal{P}}, \mathbf{T}_{\mathcal{P}}),$$

where $\mathbf{W}_{\mathcal{P}p}$ are observed covariates, $\boldsymbol{\gamma}_{\mathcal{P}}$ are regression parameters, and $\mathbf{T}_{\mathcal{P}}$ is the covariance matrix of the residuals.

Also, the items may have a hierarchical structure. Let $\boldsymbol{\xi}_{ks}$ be the parameters of an item k nested in a unit $s,$ $s = 1, \ldots, S.$ It is assumed that the first-level item parameters $\boldsymbol{\xi}_{ks}$ have multivariate normal distributions. Analogous to the case of the student parameters, we introduce observed covariates $\mathbf{X}_{\mathcal{I}ks}$ and $\mathbf{W}_{\mathcal{I}s},$ regression parameters $\boldsymbol{\beta}_{\mathcal{I}s}$ and $\boldsymbol{\gamma}_{\mathcal{I}},$ and covariance matrices of residuals $\boldsymbol{\Sigma}_{\mathcal{I}}$ and $\mathbf{T}_{\mathcal{I}},$ and we define the regression models

$$\boldsymbol{\xi}_{ks} \sim N(\mathbf{X}_{\mathcal{I}ks}\boldsymbol{\beta}_{\mathcal{I}s}, \boldsymbol{\Sigma}_{\mathcal{I}})$$

and

$$\boldsymbol{\beta}_{\mathcal{I}s} \sim N(\mathbf{W}_{\mathcal{I}s}\boldsymbol{\gamma}_{\mathcal{I}}, \mathbf{T}_{\mathcal{I}}).$$

Bayesian Estimation

Glas (2006) presented a comprehensive Bayesian estimation method for simultaneous estimation of the parameters for models for discrete and continuous responses for a broad class of IRT models. The method is a generalization and combination of methods proposed by Albert (1992), Johnson and Albert (1999), and Béguin and Glas (2001), and it is based on a Markov chain Monte Carlo (MCMC) computational method, that is, the Gibbs sampler (Gelfand & Smith, 1990). In the present report, we generalize the method further, using the Gibbs sampler to generate the posterior distributions. These distributions are simulated in an iterative process. To implement the Gibbs sampler, the parameter vector is divided into a number of components, and each successive component is sampled from its conditional distribution given sampled values for all other components. This sampling scheme is repeated until the sampled values form stable posterior distributions. For application of the Gibbs sampler, it is important to create a set of partial posterior distributions that are easy to sample from. Therefore, discrete observations are mapped to continuous normally distributed variables by a process called *data augmentation*.

Data Augmentation

Discrete observations are mapped to a latent continuous response z_{nk} in a number of data augmentation steps (in this section we will drop the higher-level indices p and s for convenience). After this mapping, the MCMC algorithm does not distinguish between continuous observed z_{nk} and latent responses z_{nk}.

The missing data indicators d_{nk} are treated as follows. First, since the single peaked model can be viewed as a collapsed version of the GRM, we map the observed dichotomous response onto (0, 1, 2) using

$$\begin{aligned} &P(U_{nk} = 0 | d_{nk} = 0, \eta_{nk}, c_k) \propto 1 - \Phi(\eta_{nk1}) \\ &P(U_{nk} = 1 | d_{nk} = 1, \eta_{nk}, c_k) = 1 \\ &P(U_{nk} = 2 | d_{nk} = 0, \eta_{nk}, c_k) \propto \Phi(\eta_{nk2}). \end{aligned} \quad (7)$$

Second, for the draws u_{nk} we assume the GRM. The data augmentation scheme for the GRM was developed by Johnson and Albert (1999). We first broaden the definition of the item parameters with $b_{k0} = -\infty$ and $b_{km} = \infty$, so we have $\eta_{nk0} = -\infty$ and $\eta_{km} = \infty$. Then simulation is based on the posterior

$$p(z_{nk} \,|\, u_{nk}, \eta_{nk}) \propto \prod_{j=1}^{m} \phi(z_{nk}; \eta_{nkj}, 1) u_{nkj} \left[\mathbf{I}(\eta_{nk(j-1)} < z_{nk} \leqslant \eta_{nkj}) \right], \tag{8}$$

where $\mathbf{I}(.)$ is an indicator function assuming the value one if its argument is true, and zero otherwise. So the factor $u_{nkj} \left[\mathbf{I}(\eta_{nk(j-1)} < z_{nk} \leqslant \eta_{nkj}) \right]$ is positive only if $u_{nkj} = 1$ and $\eta_{nk(j-1)} < z_{nk} \leqslant \eta_{nkj}$.

Posterior Simulation

The posterior distribution of all parameters and augmented data given the observations is given by

$$\begin{aligned} p(\boldsymbol{\xi}, \boldsymbol{\theta}, \tilde{\mathbf{z}}, \mathbf{u}, \boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{d}, \mathbf{z}) = {} & p(\mathbf{z}, \mathbf{u} \mid \mathbf{d}; \boldsymbol{\xi}, \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \mathbf{X}_{\mathcal{P}}, \boldsymbol{\beta}_{\mathcal{P}}, \boldsymbol{\Sigma}_{\mathcal{P}})\, p(\boldsymbol{\xi} \mid \mathbf{X}_{\mathcal{I}}, \boldsymbol{\beta}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}}) \\ & p(\boldsymbol{\beta}_{\mathcal{P}} \mid \mathbf{W}_{\mathcal{P}}, \boldsymbol{\gamma}_{\mathcal{P}}, \mathbf{T}_{\mathcal{P}}, \boldsymbol{\Sigma}_{\mathcal{P}})\, p(\boldsymbol{\gamma}_{\mathcal{P}} \mid \mathbf{T}_{\mathcal{P}})\, p(\boldsymbol{\Sigma}_{\mathcal{P}})\, p(\mathbf{T}_{\mathcal{P}}) \\ & p(\boldsymbol{\beta}_{\mathcal{I}} \mid \mathbf{W}_{\mathcal{I}}, \boldsymbol{\gamma}_{\mathcal{I}}, \mathbf{T}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}})\, p(\boldsymbol{\gamma}_{\mathcal{I}} \mid \mathbf{T}_{\mathcal{I}})\, p(\boldsymbol{\Sigma}_{\mathcal{I}})\, p(\mathbf{T}_{\mathcal{I}}), \end{aligned}$$

where $\mathbf{z}$ are the observed responses and $\tilde{\mathbf{z}}$ are the augmented latent responses.

Glas (2006) described a general procedure for generating samples from posteriors for Bayesian estimation of multidimensional IRT models for discrete and continuous responses. The first three steps needed for the present application follow directly from this general procedure:

1. Draw $\mathbf{u}$ and $\mathbf{z}$ conditional on $\boldsymbol{\theta}, \boldsymbol{\xi},$ and $\mathbf{d}$,

2. Draw $\boldsymbol{\theta}$ conditional on $\mathbf{z}, \boldsymbol{\xi}, \boldsymbol{\Sigma}_{\mathcal{P}},$ and $\boldsymbol{\mu}_{\mathcal{P}} \mathbf{X}_{\mathcal{P}}, \boldsymbol{\beta}_{\mathcal{P}}$,

3. Draw $\boldsymbol{\xi}$ conditional on $\mathbf{z}$ and $\boldsymbol{\theta}, \boldsymbol{\Sigma}_{\mathcal{I}},$ $\boldsymbol{\mu}_{\mathcal{I}}, \mathbf{u},$ and $\mathbf{d}$,

with $\boldsymbol{\mu}_{\mathcal{P}} = \mathbf{X}_{\mathcal{P}} \boldsymbol{\beta}_{\mathcal{P}}$ and $\boldsymbol{\mu}_{\mathcal{I}} = \mathbf{X}_{\mathcal{I}} \boldsymbol{\beta}_{\mathcal{I}}$.

The steps needed for the estimation of the regression coefficients and the covariance matrices of the residuals are treated in the next section. The procedure that will be presented is a multivariate extension of the procedure for MCMC estimation of the multilevel model presented by Fox and Glas (2001, 2002, 2003).

MCMC Steps for the Regression Parameters and Covariance Matrices

In this section, the models for the item and person parameters will be treated simultaneously. Therefore, we introduce parameters λ_{ijq} that may either be the parameters θ_{npq}, for persons n nested in units p, and for $q = 1, \ldots, Q,$ where Q is the dimensionality of the latent ability space, or ξ_{ksq}, for $q = 1, \ldots, Q,$ where Q is the number of item parameters of item k nested in unit s. We impose a two-level regression model on the latent variables λ_{ijq}, that is,

$$\lambda_{ijq} = \sum_{p=1}^{P} \beta_{jpq} x_{ijp} + \varepsilon_{ijq}$$

and

$$\beta_{jpq} = \sum_{s=1}^{S} \gamma_{pqs} w_{jspq} + \upsilon_{jpq}.$$

It will be assumed that x_{ij1} and w_{j1pq} are equal to one. The error terms have distributions

$$\boldsymbol{\varepsilon}_{ij} \sim N(\mathbf{0},\boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma}$ is a $Q \times Q$ covariance matrix and

$$\boldsymbol{\upsilon}_j \sim N(\mathbf{0},\mathbf{T}),$$

where $\mathbf{T}$ is a $PQ \times PQ$ covariance matrix. Both $\mathbf{T}$ and $\boldsymbol{\Sigma}$ are not restricted to be diagonal.

The priors of all covariance matrices are noninformative inverse Wishart distributions (see, for instance, Box & Tiao, 1973). This leads to the following four MCMC steps:

Step 1

Sample $\boldsymbol{\beta}_j$. Define

- $\boldsymbol{\beta}_j$ as a PQ-dimensional vector of the elements β_{jpq},
- $\boldsymbol{\lambda}_j$ as an N_jQ-dimensional vector of the elements λ_{ijq}, where N_j is the number of observations on the j,
- $\boldsymbol{\gamma}$ as a PQS-dimensional vector of the elements γ_{pqs},
- $\mathbf{W}_{jpq}$ as an S-dimensional vector of the elements w_{jspq} and $\mathbf{W}_j = \{\mathbf{W}_{jpq}\} \otimes \mathbf{I}_{PQ}$ (note that $\mathbf{W}_j$ is a $PQ \times PQS$ matrix),
- Define $\mathbf{X}_j^* = \{\mathbf{X}_j\} \otimes \mathbf{I}_Q$, with $\mathbf{X}_j$ a matrix of the elements $\{\mathbf{x}_{ijp}\}$ (note that $\mathbf{X}_j$ is an $N_jQ \times PQ$ matrix).

Given all other parameters, the conditional distribution of $\boldsymbol{\beta}_j$ is normal; that is,

$$\boldsymbol{\beta}_j \mid \boldsymbol{\lambda}_j, \boldsymbol{\Sigma}, \mathbf{T}, \boldsymbol{\gamma}, \mathbf{W}_j, \mathbf{X}_j \sim N\left(\boldsymbol{\Phi}\left[\mathbf{X}_j^{*\prime}\boldsymbol{\lambda}_j + \mathbf{T}^{-1}\mathbf{W}_j\boldsymbol{\gamma}\right] \ , \ \boldsymbol{\Phi} \right),$$

with $\boldsymbol{\Phi} = (\mathbf{X}_j^{*\prime}\mathbf{X}_j^* + \mathbf{T}^{-1})^{-1}$.

Step 2

Sample $\boldsymbol{\Sigma}$. Define the matrix of residuals $\mathbf{S} = \sum_j (\boldsymbol{\lambda}_{ij} - \mathbf{B}_j\mathbf{X}_{ij})(\boldsymbol{\lambda}_{ij} - \mathbf{B}_j\mathbf{X}_{ij})'$, with

$\boldsymbol{\lambda}_{ij}$ a Q-vector of the elements λ_{ijq},

$\mathbf{X}_{ij}$ a P-dimensional vector of the elements x_{ijp},

$\mathbf{B}_j$ a $Q \times P$ matrix of the elements β_{jpq}.

Then the conditional distribution of $\boldsymbol{\Sigma}$ is inverse Wishart:

$$\boldsymbol{\Sigma} \mid \boldsymbol{\lambda}, \mathbf{X}, \boldsymbol{\beta} \sim \text{Inv-W}(N, \mathbf{S}^{-1}).$$

Step 3

Sample $\boldsymbol{\gamma}$ from

$$\boldsymbol{\gamma} \mid \mathbf{W}, \mathbf{T}, \boldsymbol{\beta} \sim N\left(\boldsymbol{\Psi}\sum_j \mathbf{W}_j'\mathbf{T}^{-1}\boldsymbol{\beta}_j \ , \ \boldsymbol{\Psi} \right),$$

where $\boldsymbol{\Psi} = \sum_j \mathbf{W}_j'\mathbf{T}^{-1}\mathbf{W}_j{}^{-1}$.

Step 4

Sample $\mathbf{T}$. Define the matrix of residuals $\mathbf{S} = \frac{1}{J}\sum_j(\beta_j - \mathbf{W}_j\gamma)(\beta_j - \mathbf{W}_j\gamma)'$. Then the conditional distribution of $\mathbf{T}$ is inverse Wishart:

$$\mathbf{T} \mid \mathbf{B}, \mathbf{W}, \gamma \sim \text{Inv-W}(J, \mathbf{S}^{-1}).$$

An Empirical Example

Data from the Dutch central examinations in secondary education were used to show how the methods outlined above may be used in practice. The data were part of the data on pre-university students who took their final examination in the school year 1994–1995. The students chose an examination package that consisted of different examination topics. For our illustration we chose students choosing at least four of seven topics: Dutch, English, and German Language; History; Mathematics; General Economy; and Business Economy. The sample consisted of 445 students. The examination scores were on a scale from 0 to 10, with two significant digits after the decimal point.

Glas and Korobko (2005) and Glas (2006) fitted an IRT factor model with four dimensions to these data. The first three dimensions pertained to an IRT model for the grades given in (1) and (2). The dimensions were identified as Language, Science, and Economy ability, respectively. The fourth dimension pertained to the choice model given in (4). This fourth dimension correlated highly positively with the other three dimensions and was, therefore, identified as an overall ability dimension.

The first research question tackled in this report was how much of the variance in the latent person parameters was attributable to the schools. Therefore, the MCMC analysis of the previous report was redone with a two-level model (without covariates) for the ability parameters. That is, the overall covariance matrix was partitioned into a within-schools covariance matrix $\mathbf{\Sigma}_{\mathcal{P}}$ and a between-schools covariance matrix $\mathbf{T}_{\mathcal{P}}$. The Gibbs sampler was run using 20,000 iterations. The results are shown in Table 1.

TABLE 1
Bayesian estimates of parameters of examination topics[a]

Topic	a_{k1}	a_{k2}	a_{k3}	b_k	$\bar{b}_k$	$Se(a_{k1})$	$Se(a_{k2})$	$Se(a_{k3})$	$Se(b_k)$	$Se(\bar{b}_k)$
Dutch	0.20	0.16	0.60	−6.21	—	.150	.295	.299	.061	
German	1.00*	0.00*	0.00*	−6.52	−0.69				.072	.129
English	1.08	0.00*	0.00*	−6.33	—	.122			.064	
History	0.45	0.12	0.60	−6.81	−0.22	.159	.277	.277	.069	.132
Mathematics	0.00*	1.00*	0.00*	−5.91	0.05				.077	.111
Gen. Economy	0.00*	0.00*	1.00*	−6.22	−0.44				.072	.111
Bus. Economy	0.00*	0.35	0.99	−6.19	−0.22		.277	.273	.060	.114
Covariance Matrix $\mathbf{\Sigma}_{\mathcal{P}}$						$Se(\sigma_{*1})$	$Se(\sigma_{*2})$	$Se(\sigma_{*3})$	$Se(\sigma_{*4})$	
Language	0.711					.037				
Science	0.226	0.610				.043	.035			
Economy	0.297	0.456	0.530			.044	.045	.038		
Choice	0.026	0.634	0.425	1.044		.069	.065	.061	.043	
Covariance Matrix $\mathbf{T}_{\mathcal{P}}$						$Se(\tau_{*1})$	$Se(\tau_{*2})$	$Se(\tau_{*3})$	$Se(\tau_{*4})$	
Language	0.098					.065				
Science	0.008	0.006				.081	.034			
Economy	0.049	0.020	0.102			.084	.041	.136		
Choice	0.003	0.018	0.049	0.077		.124	.132	.128	.088	
Correlation Matrix $\mathcal{R}(\mathbf{\Sigma}_{\mathcal{P}})$						Correlation Matrix $\mathcal{R}(\mathbf{T}_{\mathcal{P}})$				
Language	1.000					1.000				
Science	0.343	1.000				0.339	1.000			
Economy	0.484	0.803	1.000			0.496	0.816	1.000		
Choice	0.031	0.795	0.571	1.000		0.030	0.826	0.555	1.000	

[a]Starred entries are fixed.

The point estimates reported are posterior expectations (expected a posteriori [EAP]) and posterior standard deviations (PSD). Note that the choice dimension has significant positive correlations with all proficiency dimensions. The correlation with the Science dimension was highest. For the choice dimension, displaying the factor loadings is little informative, since they are all equal to one. Therefore, the average of the two subject parameters, that is, $\bar{b}_k = (b_{k1} - b_{k2})/2$ are displayed for all subjects in the last column labeled $\bar{b}_k$. The parameters $\bar{b}_k$ can be seen as an estimate of the location of the subject on this fourth proficiency dimension. Note that the parameters for Dutch and English cannot be estimated because these two examination subjects are obligatory; therefore, all the choice variables d_{nk} for these examination subjects are structurally equal to one, and the parameters related to these subjects cannot be estimated. For further interpretation of the mean parameters $\bar{b}_k$, refer to Glas (2006). From comparing the estimates of the item parameters given in Table 3 of Glas (2006) and Table 1 of the present report, it can be seen that there was little difference between the estimates.

The within-schools covariance matrix $\mathbf{\Sigma}_{\mathcal{P}}$ and the between-schools covariance matrix $\mathbf{T}_{\mathcal{P}}$, and the associated correlation matrices $\mathcal{R}(\mathbf{\Sigma}_{\mathcal{P}})$ and $\mathcal{R}(\mathbf{T}_{\mathcal{P}})$, are given in the last panels of Table 1. The question regarding the proportion of variance in the latent person parameters attributable to the schools could, in principle, be addressed by using these variance estimates to compute intraclass correlation coefficients (ICCs; see, for instance, Bryk and Raudenbush, 1992). However, to obtain a measure of the credibility of the ICCs, it is more convenient to sample their values during the MCMC procedure and to compute their EAPs and PSDs. The ICCs were computed as the variance ratio

$$\rho = \frac{\tau^2}{\sigma^2 + \tau^2},$$

where σ^2 and τ^2 are the appropriate diagonal elements from $\mathbf{\Sigma}_{\mathcal{P}}$ and $\mathbf{T}_{\mathcal{P}}$, respectively. The posterior means and standard deviations of the six ICCs are shown in Table 2. Note that the ICCs for the Language and Economy dimensions exceeded 10%. On the other hand, the ICC for the Science dimension was very close to zero.

TABLE 2
Bayesian estimates of intraclass correlations ρ

Topic	EAP	PSD
Language	.120	.030
Science	.010	.029
Economy	.159	.034
Choice	.071	.044

The second research question concerned the proportion of variance attributable to gender. To answer this question, we carried out a second analysis with gender as a predictor for each of the four ability dimensions. As above, the question was addressed for all dimensions. The proportion of explained variance was computed as

$$\delta = \frac{\sigma^2_{\text{Model 0}} - \sigma^2_{\text{Model 1}}}{\sigma^2_{\text{Model 0}}},$$

where $\sigma^2_{\text{Model 0}}$ and $\sigma^2_{\text{Model 1}}$ are the EAPs of the appropriate diagonal elements of $\mathbf{\Sigma}_{\mathcal{P}}$ obtained in the analysis without and with gender as a covariate, respectively. The results are shown in Table 3. Note that an estimate of the reliability of the indices is lacking. The reason is that the indices are computed from two separate analyses and not sampled in a single analysis. The computation of a measure for the reliability of the estimate of δ remains a point for further research. The results are displayed in the second column of Table 3.

TABLE 3
Bayesian estimates of gender effect β and proportion variance explained δ

Topic	EAP δ	EAP β	PSD β
Language	.082	.121	.042
Science	.077	−.008	.042
Economy	.006	.001	.043
Choice	.032	−.041	.042

Note that the proportions of variance explained by gender were highest for the Language and Science dimensions. Further, the EAP estimates and the PSDs of the regression coefficient for gender are displayed in the last two columns. Male gender was coded zero; female gender was coded one. So a positive value of β reflects a higher ability level for the female students, while a negative value indicates the opposite. Note that the average for the Language dimension was higher for the female students, while their average on the overall ability dimension (the choice dimension) was lower.

References

Albert, J. H. (1992). Bayesian estimation of normal ogive item response functions using Gibbs sampling. *Journal of Educational Statistics, 17,* 251–269.

Béguin, A. A., & Glas, C. A. W. (2001). MCMC estimation and some fit analysis of multidimensional IRT models. *Psychometrika, 66,* 541–562.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Reading, MA: Addison-Wesley.

Box, G., & Tiao, G. (1973). *Bayesian inference in statistical analysis*. Reading, MA: Addison-Wesley.

Bradlow, E. T., & Thomas, N. (1998). Item response theory models applied to data allowing examinee choice. *Journal of Educational and Behavioral Statistics, 23,* 236–243.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models.* Newbury Park: Sage Publications.

Fox, J.-P., & Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika, 66*, 271–288.

Fox, J.-P., & Glas, C. A. W. (2002). Modelling measurement error in structural multilevel models. In G. A. Marcoulides and I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 245–269). Mahwah, NJ: Laurence Erlbaum.

Fox, J.-P., and C. A. W. Glas (2003), Bayesian modeling of measurement error in predictor variables using item response theory. *Psychometrika, 68*, 169–191.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85,* 398–409.

Glas, C. A. W. (2006). *Bayesian estimation methods for multidimensional models for discrete and continuous responses* (LSAC Research Report 06-05). Newtown: PA, Law School Admission Council, Inc.

Glas, C. A. W., & Korobko, O. (2005). *Likelihood-based statistics for validating continuous response models* (LSAC Research Report 05-03). Newtown: PA, Law School Admission Council, Inc.

Holman, R. & Glas, C. A. W. (2005). Modelling non-ignorable missing data mechanisms with item response theory models. *British Journal of Mathematical and Statistical Psychology, 58,* 1–17.

Johnson, V. E., & Albert, J.H. (1999). *Ordinal data modeling.* New York, NJ: Springer.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

Moustaki, I., & Knott, M. (2000). Weighting for item non-response in attitude scales by using latent variable models with covariates. *Journal of the Royal Statistical Society, A, 163,* 445–459.

Moustaki, I., & O'Muircheartaigh, C. (2000). A one dimensional latent trait model to infer attitude from nonresponse for nominal data. *Statistica, 60*(2), 259–276.

O'Muircheartaigh, C., & Moustaki, I. (1999). Symmetric pattern models: A latent variable approach to item non-response in attitude scales. *Journal of the Royal Statistical Society, A, 162,* 177–194.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63,* 581–592.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph, 17.*

Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology, 43,* 39–55.