LSAC RESEARCH REPORT SERIES

Alternative Approaches to Updating Item Parameter Estimates in Tests With Item Cloning

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

Law School Admission Council
Computerized Testing Report 03-01
September 2006

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 and Canada. It was founded in 1947 to coordinate, facilitate, and enhance the law school admission process. The organization also provides programs and services related to legal education. 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 included in the voting membership of the Council.

Table of Contents

Executive Summary

All item-cloning techniques are based on a formal description of a set of *parent items* and an algorithm to derive a larger set of operational items from them. By using such techniques, item pools for computerized adaptive testing can be created more cost effectively and item presentation can be enhanced by using more uniform item formats.

The item response theory (IRT) mathematical model is usually applied in computerized adaptive testing. In IRT, several statistics, called *item parameters*, are used to describe individual test items. One of the issues to be addressed in the application of item cloning techniques is the extent to which the item parameters will vary between parent and cloned items. Previous researchers have proposed a mathematical model, referred to in this paper as the item cloning model (ICM), to deal with possible variability of the item parameters introduced by item cloning.

The current study compares two procedures for updating the parameter estimates for an item bank consisting of cloned items calibrated under the ICM. Results from the simulation studies indicated that the first procedure tended to have a minor gain in precision in the parameter estimates. The second method was 8% faster, not a large enough gain to offset loss of precision.

Abstract

Item cloning techniques can greatly reduce the cost of item writing and enhance the flexibility of item presentation. To deal with the possible variability of the item parameters caused by item cloning, Glas and van der Linden (in press, 2006) proposed a multilevel item response model where it is assumed that the item parameters of a 3-parameter logistic (3PL) model or a 3-parameter normal ogive (3PNO) model are sampled from a multivariate normal distribution associated with a parent item. The model is referred to as the item cloning model (ICM). For the situation where each cloned item is presented to a substantial number of respondents, Glas and van der Linden (2006) proposed a Bayesian procedure for parameter estimation using a Markov chain Monte Carlo (MCMC) method (the Gibbs sampler). Two procedures for updating the parameter estimates in the ICM are compared. In the first procedure, the MCMC procedure is run on the combined original and new data set. In the second procedure, the estimates obtained on the original data set are used as priors in an MCMC run using the new data only. Results of simulation studies indicated that the second procedure tended to lead to some loss of precision in the parameter estimates. However, in the simulation studies presented here, this loss was limited. On the other hand, the gain in computation time for the second method was not substantial either.

Introduction

Item cloning is based on a formal description of a set of *parent items* and an algorithm to derive a larger set of operational items from them. These parent items have been known as *item forms*, *item templates*, or *item shells*, whereas the items generated from them are now widely known as *item clones*. Comprehensive reviews of the technology of item cloning are given in Bejar (1993) and Roid and Haladyna (1982).

Recently, Glas and van der Linden (in press, 2006) proposed a multilevel item response (IRT) model where it is assumed that the item parameters of a 3PL model are sampled from a multivariate normal distribution associated with a parent item. The model is fully Bayesian in the sense that (informative) priors are formulated for all hyperparameters describing the distributions of the item parameters within the populations. The numerical procedure used to calculate the estimates was a Markov chain Monte Carlo (MCMC) simulation (Gibbs sampler). In this paper, two procedures for updating item parameter estimates in the item cloning model (ICM) are compared. In the first procedure, the new incoming data are added to the already available old data, and the MCMC procedure is run again to produce new posterior distributions of the parameters. In the second procedure, the posterior distributions generated using the old data will be entered as prior distributions in an MCMC procedure for generating the posterior distributions of the item parameters given the new data. The motivation for considering the second procedure is that the first one is more time consuming. Therefore, the objective of the study is to evaluate whether the estimates delivered in the second procedure are also acceptable.

This paper is organized as follows. First, the ICM will be presented, and the MCMC procedure for the parameters will be sketched. Second, an approximate model that incorporates information from previous calibration runs as prior information will be introduced, and the MCMC procedure for estimating this model will be sketched. Finally, a number of simulation studies comparing the two ICM approaches will be presented, and some conclusions will be drawn.

Model

Consider a set of item populations $p = 1, ..., P$ of size $K_1, ..., K_P$, respectively. The item clones in population p will be labeled $i_p = 1, ..., K_p$. It proves convenient to introduce sampling design variables d_{ni_p}, which assume a value equal to one if student n responded to item i_p and zero otherwise. Let X_{ni_p} be the response variable for student n and item clone i_p. If $d_{ni_p} = 1$, X_{ni_p} attains the value one for a correct response and a value zero for an incorrect response. If $d_{ni_p} = 0$, X_{ni_p} attains an arbitrary value r ($r \neq 0$; $r \neq 1$). Notice that with this definition the design variables are completely determined by the response variables; they are only introduced here for mathematical convenience.

First-Level Model

The first-level model is the 3PNO model, which describes the probability of a correct response as

$$p(X_{ni_p} = 1 \mid d_{ni_p} = 1, \theta_n, a_{i_p}, b_{i_p}, c_{i_p}) = c_{i_p} + (1 - c_{i_p})\Phi(a_{i_p}\theta_n - b_{i_p}), \tag{1}$$

where a_{i_p}, b_{i_p}, and c_{i_p} are item parameters, θ_n is an examinee parameter, and $\Phi(.)$ is the cumulative normal distribution function. The parameterization of the model in (1) is slightly different from the usual parameterization for the logistic and normal-ogive models, which is $a_{i_p}(\theta - b_{i_p})$. The only motivation for this choice is to simplify the presentation below.

The reason for considering the 3PNO model rather that the 3PL model is that the former appears to be more tractable in an MCMC framework. However, as is well known, for an appropriately chosen scale factor, both the parameter estimates of the models are numerically nearly indistinguishable, and either model is expected to fit only if the other does.

Second-Level Model

The values of the item parameters (a_{i_p}, b_{i_p}, and c_{i_p}) in (1) are considered as realizations of a random vector. It is assumed that

$$\boldsymbol{\xi}_{i_p} = (a_{i_p}, b_{i_p}, \text{logit } c_{i_p}), \tag{2}$$

has a multivariate normal distribution; that is,

$$\boldsymbol{\xi}_{i_p} \sim N(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p), \tag{3}$$

where $\boldsymbol{\mu}_p$ is the vector with the mean values of the item parameters for population p and $\boldsymbol{\Sigma}_p$ their covariance matrix. The hyperparameters ($\boldsymbol{\mu}_p$, $\boldsymbol{\Sigma}_p$) are allowed to vary across the populations of items. Finally, it will be assumed that θ_n has a standard normal distribution

$$\theta_n \sim N(0, 1). \tag{4}$$

Prior Distributions

It proves convenient to assume separate prior distributions for all populations p. Using a normal-inverse-Wishart distribution (see, for instance, Box & Tiao, 1973, or Gelman, Carlin, Stern & Rubin, 1995), the prior for ($\boldsymbol{\mu}_p$, $\boldsymbol{\Sigma}_p$) is

$$\boldsymbol{\Sigma}_p \sim Inv - Wishart_{v0}(\boldsymbol{\Sigma}_0)$$

and

$$\boldsymbol{\mu}_p \mid \boldsymbol{\Sigma}_p \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_p / \kappa_0),$$

where $\mathbf{\Sigma}_0$ and v_0 are the scale matrix and degrees of freedom for the prior on $\mathbf{\Sigma}_p$, respectively, and $\boldsymbol{\mu}_0$ and κ_0 are the mean and weight for the prior on $\boldsymbol{\mu}_p$, respectively. The weights κ_0 and v_0 express the information in the prior distribution.

Likelihood Function

The response vector of examinee n is denoted as $\mathbf{x}_n = (x_{ni_1}, ..., x_{ni_p}, ..., x_{ni_P})$. Using the assumptions of (1) independence between examinees, (2) independence between items and examinees, and (3) local independence within examinees, the likelihood function associated with response data $\mathbf{x}$ and test administration design $\mathbf{d}$ can be written as

$$p(\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{x}, \mathbf{d}) = \prod_n p(\mathbf{x}_n \mid \mathbf{d}_n, \theta_n, \boldsymbol{\xi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$= \prod_n \prod_p \prod_{i_p} p(x_{ni_p} \mid d_{ni_p}, \theta_n, \boldsymbol{\xi}_{i_p}) p(\theta_n)$$

$$\prod_p \prod_{i_p} p(\boldsymbol{\xi}_{i_p} \mid \boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p). \tag{5}$$

The convention will be followed that $p(x_{ni_p} = r \mid d_{ni_p} = 0, \theta_n, a_{i_p}, b_{i_p}, c_{i_p}) = 1$.

Bayesian Estimation

An MCMC procedure will be used to sample from the posterior distribution. Only the essential steps will be outlined here; for details consult, Glas and van der Linden (2006). Following Albert (1992) and Béguin and Glas (2001), two data augmentation schemes are used to create tractable posterior distributions. First, a binary variable W_{ni_p} is introduced with a conditional distribution given by

$$P(W_{ni_p} = 1 \mid X_{ni_p} = 1, \eta_{ni_p}, c_{i_p}) \propto \Phi(\eta_{ni_p})$$

$$P(W_{ni_p} = 0 \mid X_{ni_p} = 1, \eta_{ni_p}, c_{i_p}) \propto c_{i_p}(1 - \Phi(\eta_{ni_p}))$$

$$P(W_{ni_p} = 1 \mid X_{ni_p} = 0, \eta_{ni_p}, c_{i_p}) = 0$$

$$P(W_{ni_p} = 0 \mid X_{ni_p} = 0, \eta_{ni_p}, c_{i_p}) = 1, \tag{6}$$

where $\eta_{ni_p} = a_{i_p}\theta - b_{i_p}$. Second, the data are augmented with latent data Z_{ni_p}, which are independent and normally distributed with mean η_{ni_p} and standard deviation equal to one.

The aim of the procedure is to simulate samples from the joint posterior distribution given by

$$p(\boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{z}, \mathbf{w} \mid \mathbf{x}) \propto p(\mathbf{z}, \mathbf{w} \mid \mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) p(\boldsymbol{\xi} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0). \tag{7}$$

The samples are generated with a Gibbs sampler consisting of four steps.

Step 1. Draw from $p(\mathbf{z}, \mathbf{w} \mid \mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\theta})$

Step 2. Draw from $p(\boldsymbol{\theta} \mid \mathbf{z}, \boldsymbol{\xi})$

Step 3. Draw from $p(\boldsymbol{\xi}_{i_p} \mid \boldsymbol{\theta}, \mathbf{z}_{i_p}, \boldsymbol{\mu}_{i_p}, \boldsymbol{\Sigma}_{i_p})$

Step 4. Draw from $p(\mu_p, \Sigma_p \mid \boldsymbol{\xi}, \boldsymbol{\theta}, \mathbf{z}, \mathbf{x})$.

Multiple MCMC chains can be started from different points to evaluate convergence by comparing the between- and within-sequence variance. Another approach is to generate a single MCMC chain and to evaluate convergence by dividing the chain into subchains and comparing between- and within-subchain variance. For these and other technical details, see Gelman, et al. (1995).

The Approximate Model

The approximation is based on the idea that the item clone parameters $\boldsymbol{\xi}_{i_p}$ can be estimated with sufficient precision in the first stage (say Phase I) using the ICM (including the collateral information on their parentage; that is, their nesting under p), and that these estimates can serve as (highly informative) priors in the second stage (Phase II). It is assumed that the nested structure of the item clones is sufficiently reflected in the priors, so the nested structure is not taken into account in the probability model used as approximation. Therefore, the posterior distribution is

$$p(\boldsymbol{\theta}, \boldsymbol{\xi} \mid \mathbf{x}, \mathbf{d}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \prod_n \prod_p \prod_{i_p} p(x_{ni_p} \mid d_{ni_p}, \theta_n, \boldsymbol{\xi}_{i_p}) p(\theta_n)$$

$$\prod_p \prod_{i_p} p(\boldsymbol{\xi}_{i_p} \mid \boldsymbol{\mu}_{i_p}, \boldsymbol{\Sigma}_p / n_{i_p}). \tag{8}$$

In this model, $p(\boldsymbol{\xi}_{i_p} \mid \boldsymbol{\mu}_{i_p}, \boldsymbol{\Sigma}_p / n_{i_p})$ is an informative prior that summarizes the information of Phase 1. Note that $p(\boldsymbol{\xi}_{i_p} \mid \boldsymbol{\mu}_{i_p}, \boldsymbol{\Sigma}_p / n_{i_p})$ has a normal distribution, where the mean $\boldsymbol{\mu}_{i_p}$ is equal to the posterior expectation of $\boldsymbol{\xi}_{i_p}$ obtained in Phase I and the covariance matrix reflects the precision of these estimates. That is, the covariance matrix is equal to the posterior expectation obtained in Phase I, $\boldsymbol{\Sigma}_p$, divided by the number of students responding to item i_p in Phase I, say n_{i_p}.

Estimation is done by an MCMC procedure analogous to the one above, with the first two steps unchanged and the fourth step deleted. The third step is modified as follows.

Step 3. The vector of random item parameters $\boldsymbol{\xi}_{i_p}$ is partitioned into $\boldsymbol{\delta} = (\boldsymbol{\delta}_{i_p}) = (a_{1_1}, b_{1_1}, ..., a_{i_p}, b_{i_p}, ...)$ and $\mathbf{c} = (c_{1_1}, ..., c_{i_p}, ...)$. Hence, their conditional posterior density factors $p(\boldsymbol{\xi}_{i_p} \mid \boldsymbol{\theta}, \mathbf{z}_{i_p}, \boldsymbol{\mu}_{i_p}, \boldsymbol{\Sigma}_{i_p}) = p(\text{logit } c_{i_p} \mid \boldsymbol{\delta}_{i_p}, \boldsymbol{\theta}, \mathbf{z}_{i_p}, \boldsymbol{\mu}_{c|\delta}, \boldsymbol{\Sigma}_{c|\delta}) p(\boldsymbol{\delta}_{i_p} \mid \boldsymbol{\theta}, \mathbf{z}_{i_p}, \boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)$, where $\boldsymbol{\mu}_{c|\delta}$ and $\boldsymbol{\Sigma}_{c|\delta}$ are the expectation and variance of logit c_{i_p} conditional on $\boldsymbol{\delta}_{i_p}$. Then the following two substeps are made:

a) The value of δ_{i_p} is drawn from the conditional posterior distribution of the parameters of $\boldsymbol{\delta}$ given $\boldsymbol{\theta}$, $\mathbf{z}_{i_p}$, $\boldsymbol{\mu}_{i_p}$, and $\boldsymbol{\Sigma}_p / n_{i_p}$. The distribution is derived as follows: Parameters $\boldsymbol{\delta}_{i_p}$ can be viewed as coefficients of the regression of $\mathbf{z}_{i_p} = (z_{ni_p})$, on $\mathbf{X} = (\boldsymbol{\theta}, \mathbf{-1})$, with $\mathbf{-1}$ being a column vector with entries –1. So we have $\mathbf{z}_{i_p} = \mathbf{X}\boldsymbol{\delta}_{i_p} + \varepsilon_{i_p}$. Only examinees responding to item i_p are considered here. Further, $\boldsymbol{\delta}_{i_p}$ has a normal prior with mean $\boldsymbol{\mu}_p$ and variance $\boldsymbol{\Sigma}_p$. Define $\hat{\boldsymbol{\delta}}_{i_p} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{z}_{i_p}$, define $\mathbf{d} = \mathbf{X}^t\mathbf{X}\hat{\boldsymbol{\delta}}_{i_p} + n_{i_p} \boldsymbol{\Sigma}_p^{-1} \boldsymbol{\mu}_p$ and define

$\mathbf{D} = (\mathbf{X}^t\mathbf{X} + n_{i_p} \boldsymbol{\Sigma}_p^{-1})^{-1}$. Then a well-known result from Bayesian regression analysis (see, for instance, Box & Tiao, 1973) is that

$$\boldsymbol{\delta}_{i_p} \mid \boldsymbol{\theta}, \mathbf{z}_{i_p}, \mathbf{X}, \boldsymbol{\mu}_p \boldsymbol{\Sigma}_p / n_{i_p} \sim N(\mathbf{Dd}, \mathbf{D}). \tag{9}$$

b) The value of c_{i_p} is sampled from the conditional posterior distribution given $\boldsymbol{\delta}_{i_p}$, $\boldsymbol{\theta}$, $\mathbf{z}_{i_p}$, $\boldsymbol{\mu}_{c|\delta}$, and $\boldsymbol{\Sigma}_{c|\delta}$. Let t_{i_p} be the number of students who do not know the correct answer to item i_p and guess the response. For the probability of a correct response of a student n on item i_p given $w_{ni_p} = 0$, it thus holds that $P(Y_{ni_p} = 1 \mid W_{ni_p} = 0) = c_{i_p}$. The number of correct responses obtained by guessing, S_{i_p}, say, has a binomial distribution with parameters c_{i_p} and t_{i_p}. Since logit c_{i_p} has a normal prior with parameters $\boldsymbol{\mu}_{c|\delta}$ and $\boldsymbol{\Sigma}_{c|\delta}$, the procedure for sampling in a generalized linear model with a logit-link and a normal prior (see, Gelman et al., 1995, sects 9.9 and 10.6) can be used.

Some Numerical Examples and Conclusions

The study consisted of two parts. In the first part, a real data set was analyzed to get some idea of the within-items covariance matrix. Then, in the second part, the estimates obtained in the first stage were used in a number of simulation studies aimed at assessing the quality of parameter recovery.

The data set consisted of the responses of a sample of 4,000 students from the population participating in the 1991 central examination of French language comprehension in Secondary Education in the Netherlands. The test was a traditional paper-and-pencil test. Students were clustered in 116 schools, and it was assumed that the item parameters varied across classes. It was expected that the item-parameter variance might be high in the first example and low in the second example. The MCMC method for the ICM was run with 13,000 iterations, 3,000 of which were burn-in iterations. Below, expected a-posteriori (EAP) estimates are reported as point estimates.

The values of the prior covariance matrix $\boldsymbol{\Sigma}_0$ are shown in Table 1; they are no more than an educated guess. For instance, the negative covariance between the discrimination parameter a_{i_p} and the logit-guessing-parameter logit c_{i_p} were based on the consideration that, to obtain similar item characteristic curves (ICCs), the discrimination parameter must go down if the lower asymptote goes up. In the same manner, when the respondents are relatively proficient, lowering the item difficulty parameter can be counterbalanced by lowering the discrimination parameter. This feature accounted for the choice of a positive prior covariance between the two parameters. The prior for the parent item parameters was chosen equal to $\boldsymbol{\mu}_0$ = (1.0, 0.0,logit(0.25)). To obtain convergence in the analysis of the language comprehension data, it turned out that the parameters in the normal-inverse-Wishart prior for ($\boldsymbol{\mu}_p$, $\boldsymbol{\Sigma}_p$) had to be set equal to $v_0 = 10$ and $\kappa_0 = 10$, respectively. Since $k_p = 160$, this choice results in a slightly informative prior.

The averages of the EAP point estimates of the covariance matrices are shown in Table 1 (first three columns), together with their posterior standard deviations (last three columns). It can be seen that both the posterior variance of the item discrimination and difficulty parameters were generally lower than expected.

TABLE 1
Prior and posterior values item covariance matrix

Prior Covariance Matrix						
Parameter	a	b	logit c	a	b	logit c
a	.200					
b	.100	1.000				
logit c	–.050	.050	.100			
EAP Estimate				SE		
a	.102			.017		
b	.031	.208		.017	.033	
logit c	–.018	.010	.116	.018	.020	.039

The second part of this study was aimed at assessing the precision of the estimation procedures. The averages of the EAP estimates of the mean and covariance matrix obtained using the French language examination were used as $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$. The parent item parameters, $\boldsymbol{\mu}_p$, were drawn from a normal distribution indexed by $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$, and $\boldsymbol{\Sigma}_p$ was set equal to $\boldsymbol{\Sigma}_p$. Two simulation studies were conducted.

In the first simulation study, the data for Phase I (the initial calibration) were generated as follows. For each population p, 10 items were randomly drawn from a normal distribution with parameters $\boldsymbol{\mu}_p$ and $\boldsymbol{\Sigma}_p$. To produce realistic data, parent and random item discrimination parameters drawn below 0.5 were truncated to 0.5. The responses to the random items were generated for simulees with an ability parameter randomly drawn from a standard normal distribution. Every simulee responded to 20 random items from 20 different populations. So the total data matrix consisted of 1,000 responses. The MCMC procedure for the ICM consisted of 13,000 iterations, including 3,000 burn-in iterations. To obtain convergence, the parameters in the normal-inverse-Wishart prior had to be set equal to $v_0 = 2$ and $\kappa_0 = 2$, respectively. Since $k_p = 10$, this choice entails a quite informative prior. For Phase II, new data were added to the original data set, such that $n_p = 100$ was changed to $n_p = 400$. So the sample size rose from 1,000 to 4,000. Two MCMC procedures were run; one for the ICM and one for the approximate model. For both procedures, the number of iterations and burn-in iterations were as in Phase I.

The second simulation study had a similar set-up. However, the number of item populations was equal to 40, the number of random items per population was equal to 20, and the number of responses to each random item was 200 in Phase I and 400 in Phase II. So in this case the total number of responses was equal to 4,000 in Phase I and 8,000 in Phase II.

Some results of the two simulations are presented in Table 2 and Table 3, respectively. The results are EAPs averaged over 10 replications and all items. The rows labeled a, b and logit c relate to random item parameters; all other rows relate to item-population parameters. The columns labeled MAE give the mean absolute error of the estimates, averaged over items and replications. The columns labeled SE give the posterior standard deviation for the ICM and the approximate estimates, respectively, again averaged over items and replications.

TABLE 2
Parameter recovery $P = 20$, $k_p = 10$

				ICM		Approx.	
Phase	n_p	Parameter	True	MAE[a]	SE	MAE[a]	SE
I	200	a	1.000	0.404	0.334		
		b	0.000	0.514	0.346		
		logit c	–1.099	0.327	0.654		
		μ_a	1.000	0.311	0.199		
		μ_b	0.000	0.494	0.307		
		$\mu_{\text{logit } c}$	–1.099	0.214	0.414		
		σ_a^2	0.200	0.076	0.235		
		σ_b^2	1.000	0.289	0.684		
		$\sigma_{\text{logit } c}^2$	0.100	0.377	0.550		
		$\sigma_{a,b}$	0.100	0.089	0.289		
		$\sigma_{a,\text{logit } c}$	–0.050	0.046	0.227		
		$\sigma_{b,\text{logit } c}$	0.050	0.071	0.470		
II	800	a	1.000	0.244	0.199	0.234	0.232
		b	0.000	0.275	0.222	0.258	0.218
		logit c	–1.099	0.200	0.189	0.209	0.211
		μ_a	1.000	0.151	0.138		
		μ_b	0.000	0.233	0.210		
		$\mu_{\text{logit } c}$	–1.099	0.108	0.109		
		σ_a^2	0.200	0.032	0.099		
		σ_b^2	1.000	0.145	0.209		
		$\sigma_{\text{logit } c}^2$	0.100	0.166	0.287		
		$\sigma_{a,b}$	0.100	0.033	0.095		
		$\sigma_{a,\text{logit } c}$	–0.050	0.022	0.120		
		$\sigma_{b,\text{logit } c}$	0.050	0.039	0.130		

[a]MAE refers to mean absolute error of the estimates averaged over items and replications.

TABLE 3
Parameter recovery P = 40, k_p = 20

				ICM		Approx.	
Phase	n_p	Parameter	True	MAE[a]	SE	MAE[a]	SE
I	100	a	1.000	0.392	0.204		
		b	0.000	0.365	0.192		
		logit c	–1.099	0.306	0.294		
		μ_a	1.000	0.298	0.095		
		μ_b	0.000	0.318	0.124		
		$\mu_{\text{logit } c}$	–1.099	0.199	0.130		
		σ_a^2	0.200	0.044	0.065		
		σ_b^2	1.000	0.100	0.118		
		$\sigma_{\text{logit } c}^2$	0.100	0.011	0.050		
		$\sigma_{a,b}$	0.100	0.037	0.066		
		$\sigma_{a,\text{logit } c}$	–0.050	0.009	0.043		
		$\sigma_{b,\text{logit } c}$	0.050	0.016	0.055		
II	400	a	1.000	0.197	0.188	0.222	0.234
		b	0.000	0.110	0.099	0.127	0.131
		logit c	–1.099	0.144	0.140	0.181	0.177
		μ_a	1.000	0.155	0.176		
		μ_b	0.000	0.167	0.101		
		$\mu_{\text{logit } c}$	–1.099	0.100	0.104		
		σ_a^2	0.200	0.022	0.022		
		σ_b^2	1.000	0.045	0.044		
		$\sigma_{\text{logit } c}^2$	0.100	0.009	0.009		
		$\sigma_{a,b}$	0.100	0.019	0.022		
		$\sigma_{a,\text{logit } c}$	–0.050	0.003	0.027		
		$\sigma_{b,\text{logit } c}$	0.050	0.010	0.023		

[a]MAE refers to mean absolute error of the estimates averaged over items and replications.

Especially in the case P = 40, the estimates of the covariance matrices seemed much more precise than the estimates of the item parameters. This result, however, is explained by the fact that the covariance matrices were not varied over populations, but chosen equal to their prior values. Further inspection of the results shows that the approximate procedure leads to some loss of precision in most cases, particularly when P = 40 (Table 3). However, this loss was limited to approximately 10% of the mean absolute error of the estimates. On the other hand, the time saved was not much either: both the procedure for the ICM and the approximate procedure used 13,000 iterations, but on average, the latter procedure was only 8% faster.

References

Albert, J. H. (1992). Bayesian estimation of normal-ogive item response curves using Gibbs sampling. *Journal of Educational and Behavioral Statistics, 17,* 261–269.

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

Bejar, I. I. (1993). A generative approach to psychological and educational measurement. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 323–357). Hillsdale, NJ: Lawrence Erlbaum.

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

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London: Chapman and Hall.

Glas, C. A. W., & van der Linden, W. J. (in press). *Modeling adaptive testing with item cloning* (Computerized Testing Report 01-03). Newtown, PA: Law School Admission Council.

Glas, C. A. W., & van der Linden, W. J. (2006). *Modeling Variability in Item Parameters in Educational Measurement* (Computerized Testing Report 01-07). Newtown, PA: Law School Admission Council.

Roid, G., & Haladyna, T. (1982). *A technology for test-item writing*. New York: Academic Press.