DOCUMENT RESUME

ED 464 142 TM 033 825

| | |
|---|---|
| AUTHOR | de la Torre, Jimmy; Patz, Richard J. |
| TITLE | A Multidimensional Item Response Theory Approach to Simultaneous Ability Estimation. |
| PUB DATE | 2002-04-00 |
| NOTE | 18p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 2-4, 2002). |
| PUB TYPE | Information Analyses (070) -- Speeches/Meeting Papers (150) |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | *Ability; *Bayesian Statistics; *Estimation (Mathematics); *Item Response Theory; Simulation |
| IDENTIFIERS | *Item Parameters; Multidimensionality (Tests) |

ABSTRACT

It is not unusual for several tests measuring different abilities to be given in one test administration. A common practice is to estimate these abilities independently of each other, but this paper proposes a more efficient method of estimating these abilities that takes into account the correlational structure of the abilities. The method uses a hierarchical Bayesian approach to simultaneous estimation of abilities based on a simple structure multidimensional item response model. Whether the simultaneous estimation of abilities from different dimensions yields more accurate estimates was studied, as well as how the number of dimensions, the number of items in each dimension, and the degree of correlation between abilities affect the accuracy of these estimates. For each combination of number of abilities and number of items, item parameters were randomly drawn from a pool of 550 items from nationally standardized mathematics tests, with 1,000 examinees simulated for each experimental condition. Results show that the proposed multidimensional approach gives more general outcomes, yielding results similar to those of the unidimensional method when abilities are uncorrelated. When abilities are correlated, taking the correlation into account can result in noticeable improvements in ability estimates. (Contains 24 tables and 15 references.) (SLD)

# A Multidimensional Item Response Theory Approach to Simultaneous Ability Estimation.

Jimmy de la Torre
Richard J. Patz

# A Multidimensional Item Response Theory Approach to Simultaneous Ability Estimation[1]

Jimmy de la Torre
*University of Illinois at Urbana-Champaign*

Richard J. Patz
*CTB/McGraw-Hill*

## 1 Introduction

It is not unusual for several tests measuring different abilities to be given in one test administration. Although these tests may tap different latent abilities, the abilities are usually not independent of one another. For example, in cognitive tests such as NAEP, these abilities have high positive correlations, typically greater than 0.70 (Johnson & Carlson, 1994). However, a common practice in educational measurement is to estimate these abilities independently of each other. This paper proposes a more efficient method of estimating these abilities that takes into account the correlational structure of the abilities. The method uses a hierarchical Bayesian approach to simultaneous estimation of abilities based on a simple structure multidimensional item response theory model.

## 2 Purpose

The primary purpose of this paper to investigate whether the simultaneous estimation of abilities from different dimensions yields more accurate estimates. In addition, the paper examines how the number of dimensions, the number of items in each dimension, and the degree of correlation between abilities affect the accuracy of the estimates.



## 3 Presentation of the Model

To extend the three-parameter logistic (3PL) model (Lord, 1980) to the multidimensional context, Reckase (1996) used the following generalization:

[1]Paper presented at the Annual Meeting of the National Council on Measurement in Education, April 2002, New Orleans, LA

$$P(X_{ij} = 1|\theta_i, \alpha_j, \beta_j, \gamma_j) = \gamma_j + (1 - \gamma_j)\frac{e^{\alpha_j'\theta_i+\beta_j}}{1 + e^{\alpha_j'\theta_i+\beta_j}} \tag{1}$$

where

$P(X_{ij} = 1|\theta_i, \alpha_j, \beta_j, \gamma_j)$ is the probability of examinee $i$ responding to item $j$ correctly;

$X_{ij}$ is the response of examinee $i$ to item $j$ ($0 =$ incorrect, $1 =$ correct);

$\theta_i$ is the ability vector of the examinee;

$\alpha_j$ is the vector of item parameters related to the discrimination power of the item;

$\beta_j$ is the parameter related to the difficulty of the item;

$\gamma_j$ is the pseudo-guessing parameter of the item;

$i = 1, \ldots, I$ (the total number of examinees); and

$j = 1, \ldots, J$ (the total number of items).

For this paper, simple structure is assumed (i.e., each item measures one dimension of ability and thus $\alpha_j$ contains only one non-zero element).

The model in 1 can be reexpressed as:

$$P(X_{ij(d)} = 1|\theta_{i(d)}, \alpha_{j(d)}, \beta_{j(d)}, \gamma_{j(d)}) = \gamma_{j(d)} + (1 - \gamma_{j(d)})\frac{e^{\alpha_{j(d)}\theta_{i(d)}+\beta_{j(d)}}}{1 + e^{\alpha_{j(d)}\theta_{i(d)}+\beta_{j(d)}}} \tag{2}$$

where

$X_{ij(d)}$ is the response of examinee $i$ to the $j^{th}$ item of dimension $d$;

$\theta_{i(d)}$ is the $d^{th}$ component of the vector $\theta_i$, i.e., $\theta_i = \{\theta_{i(d)}\}$;

$d = 1, \ldots, D$ (the number of dimensions);

$j(d) = 1, \ldots, J(d)$; and

$\sum_{d=1}^{D} J(d) = J$.

Refer to figure 1 for a graphical representation of the hierarchical structure of the model.

The item response $X_{ij(d)}$ has a likelihood $P_{ij(d)}$ given by

$$\begin{aligned} P_{ij(d)} &= (P(X_{ij(d)} = 1|\theta_{i(d)}, \alpha_{j(d)}, \beta_{j(d)}, \gamma_{j(d)}))^{X_{ij(d)}} \\ &\quad (1 - P(X_{ij(d)} = 1|\theta_{i(d)}, \alpha_{j(d)}, \beta_{j(d)}, \gamma_{j(d)}))^{1-X_{ij(d)}}. \end{aligned} \tag{3}$$

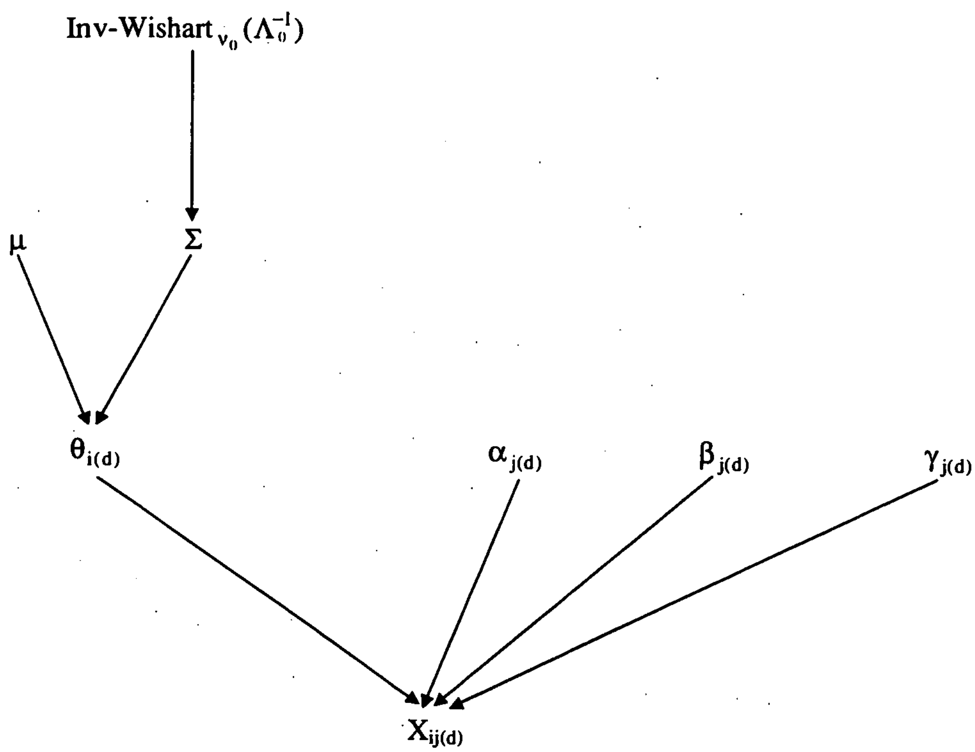


Figure 1: Graphical Representation of the Model

Let $X_{i\cdot} = \{X_{i1(1)}, \ldots, X_{iJ(1)}, \ldots, X_{i1(d)}, \ldots, X_{iJ(d)}, \ldots, X_{i1(D)}, \ldots, X_{iJ(D)}\}$ represent the response vector of examinee $i$. The corresponding likelihood of this vector is

$$P_i = \prod_{d=1}^{D} \prod_{j(d)=1}^{J(d)} P_{ij(d)}. \tag{4}$$

Finally, the likelihood of the data matrix $X$ is given by

$$P = \prod_{i=1}^{I} \prod_{d=1}^{D} \prod_{j(d)=1}^{J(d)} P_{ij(d)}. \tag{5}$$

# 4 Markov Chain Monte Carlo Estimation

The parameterization of the prior distribution of $\theta_i$ (Gelman et al., 1995) is

$$\theta_i|\Sigma \sim \text{MVN}(0, \Sigma) \tag{6}$$

$$\Sigma \sim \text{Inv-Wishart}_{\nu_0}(\Lambda_0^{-1}). \tag{7}$$

Of primary interest is the joint distribution of $\theta$ and $\Sigma$. Using the notations $X = \{X_{ij}\}$, $\theta = \{\theta_i\}$, $\alpha = \{\alpha_j\}$, $\beta = \{\beta_j\}$, and $\gamma = \{\gamma_i\}$, this joint posterior can be expressed as

$$P(\theta, \Sigma|X, \alpha, \beta, \gamma) \propto P(X|\theta, \Sigma, \alpha, \beta, \gamma)P(\theta|\Sigma)P(\Sigma). \tag{8}$$

The posterior distribution in 8 can not be evaluated in a straight-forward manner (i.e., samples cannot be drawn directly from the joint posterior distribution). Markov chain Monte Carlo (MCMC) simulation is used to draw samples iteratively from the full conditional distributions $\theta|X, \Sigma, \alpha, \beta, \gamma$ and $\Sigma|X, \theta, \alpha, \beta, \gamma$ (Casella & George, 1992; Gamerman, 1997).

For each examinee, the full conditional distribution is

$$P(\theta_i|X_i, \Sigma, \alpha, \beta, \gamma) \propto |\Sigma|^{-1/2} e^{-\frac{1}{2}\theta_i'\Sigma^{-1}\theta_i} P_i. \tag{9}$$

Although 9 is not a known distribution, samples can be drawn from this distribution indirectly by using the Metropolis-Hastings algorithm (Chib & Greenberg, 1995; Gilks et al., 1996; Tierney, 1994).

The full conditional distribution of $\Sigma$ is

$$P(\Sigma|X, \theta, \alpha, \beta, \gamma) = P(\Sigma|\theta) \propto P(\Sigma)P(\theta|\Sigma). \tag{10}$$

With the use of the prior distribution and the hyperdistributions given in 6 and 7, the full conditional posterior distribution of $\Sigma$ is an Inv-Wishart$_{\nu_I}(\Lambda_I^{-1})$, where $\nu_I = \nu_0 + I$, and $\Lambda_I = \Lambda_0 + \sum \theta_i \theta_i'$.

The full conditional distribution of $\Sigma$ is a known distribution and can be sampled directly.

For the present paper, each chain is iterated 10,000 times. The first 2,000 iterations are discarded and inference is based on the remaining 8,000 iterations.

Two methods of estimating ability are employed. The first method, which is based on the MCMC output, is called the multidimensional expected a posteriori (EAP-M) method and is computed as:

$$\tilde{\theta} = \mathrm{E}(\theta|X, \alpha, \beta, \gamma) \approx \frac{1}{8000} \sum_{t=2001}^{10000} \theta^{(t)}. \tag{11}$$

The second is based on the expected a posteriori when the correlations between abilities are assumed to be zero. Except for this assumption, this method is equivalent to the first (i.e., estimates are based also based on MCMC draws). This is called the unidimensional expected a posteriori (EAP-U) and is computed as:

$$\hat{\theta} = \mathrm{E}(\theta|X, \alpha, \beta, \gamma, \rho = 0). \tag{12}$$

Two methods are employed is estimating the underlying correlational structure between the abilities. The first method used directly estimated the correlations from the data using MCMC simulation. The covariance matrix is estimated as $\tilde{\Sigma} = \mathrm{E}(\Sigma|X, \alpha, \beta, \gamma)$. This is similar to the ability estimates, in that this is the average of the covariance matrices from the MCMC draws. The estimated covariance is standardized to obtain the correlation estimates $\tilde{\rho}$. The second estimate is based on the two-step approach where the estimate of the correlation is given by the correlation of the estimated abilities using the EAP-U method, (i.e., $\hat{R} = \mathrm{Cor}(\hat{\theta})$). For example, the correlation between $\theta_1$ and $\theta_2$ is estimated as $\hat{\rho} = \text{Cor}(\hat{\theta}_1, \hat{\theta}_2)$.

The accuracy of the ability and correlation estimates is gauged by comparing them to the generating parameters. In addition, because multiple values are available for each ability estimation method, statistical measures - Pearson correlation and root mean squared error (RMSE)- that summarize the correspondence between the estimated and the generated abilities are computed. Finally, the effectiveness of the proposed method is assessed by computing its efficiency, defined as the ratio of the average posterior variances of the ability estimates obtained using the EAP-M and the EAP-U methods.

## 5 Design of the Study

The factors investigated in this paper are: (i) the number of abilities, (ii) the number of items, and (iii) the degree of correlation between the abilities. The different number of abilities are 2 and 5, the number of items equals 10, 30 or 50 items, and the degree of correlation equals 0.00, 0.40, 0.70, and 0.90. The levels of each factor are crossed completely to yield 24 experimental conditions.

For each combination of number of abilities and number of items, item parameters are randomly drawn from a pool of 550 items that are obtained from nationally standardized mathematics tests. For each experimental condition, 1,000 examinees are drawn from $\text{MVN}(0, \Sigma_D)$,

$$\Sigma_D = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & & \vdots \\ \vdots & & \ddots & \rho \\ \rho & \dots & \rho & 1 \end{pmatrix}.$$

The responses of the examinees to the items simulated. The constraint on $\Sigma$ retains the structure of the design and does not in anyway affect the estimation process.

# 6 Results

## 6.1 Correlation estimates

Tables 1 through 6 present the correlation estimates using the algorithm that utilized the correlational structure and the two-step procedure. Although the generating value of the correlation matrix have been specified, the generated ability parameters have correlation matrix that is not identical to the generating correlational structure. Hence, comparisons must be made between the correlation estimates and the generated correlation matrix, not the generating correlation matrix.

The tables demonstrate that the correlation is well estimated by either method when there is no correlation between abilities. However, when the abilities are correlated, the correlation is underestimated by the two-step method whereas estimates based on MCMC are closer to the generated values. Increasing the number of items increases the precision of the estimates for both methods. Estimates using the two methods are not affected by increasing the number of abilities. This is to be expected for the two-step approach since it ignores the additional information contained in the correlation matrix. In general, additional precision can be expected for the MCMC estimates as more abilities are considered. However, because the large number of examinees allowed for accurate estimation of the correlations even when only two abilities, the additional information afforded by adding more abilities became negligible in the process.

Table 1: Correlation estimates for 2 dimensions and 10 items

| Method | | $\rho$ | | | |
|---|---|---|---|---|---|
| | | 0.00 | 0.40 | 0.70 | 0.90 |
| Generated | $\mathrm{Cor}(\theta_1, \theta_2)$ | 0.00 | 0.42 | 0.70 | 0.90 |
| MCMC | $\tilde{\rho}$ | -0.01 | 0.43 | 0.68 | 0.90 |
| Two-step | $\mathrm{Cor}(\hat{\theta}_1, \hat{\theta}_2)$ | -0.01 | 0.29 | 0.46 | 0.62 |

Table 2: Correlation estimates for 2 dimensions and 30 items

| Method | | $\rho$ | | | |
|---|---|---|---|---|---|
| | | 0.00 | 0.40 | 0.70 | 0.90 |
| Generated | $\mathrm{Cor}(\theta_1, \theta_2)$ | 0.05 | 0.45 | 0.68 | 0.90 |
| MCMC | $\tilde{\rho}$ | 0.06 | 0.44 | 0.69 | 0.92 |
| Two-step | $\mathrm{Cor}(\hat{\theta}_1, \hat{\theta}_2)$ | 0.05 | 0.38 | 0.59 | 0.78 |

Table 3: Correlation estimates for 2 dimensions and 50 items

| Method | | $\rho$ 0.00 | 0.40 | 0.70 | 0.90 |
|---|---|---|---|---|---|
| Generated | Cor($\theta_1, \theta_2$) | -0.06 | 0.39 | 0.70 | 0.90 |
| MCMC | $\tilde{\rho}$ | -0.06 | 0.37 | 0.69 | 0.91 |
| Two-step | Cor($\hat{\theta}_1, \hat{\theta}_2$) | -0.05 | 0.34 | 0.62 | 0.83 |

Table 4: Correlation estimates for 5 dimensions and 10 items

| Method | | $\rho$ 0.00 | 0.40 | 0.70 | 0.90 |
|---|---|---|---|---|---|
| Generated | Cor($\theta_1, \theta_2$) | -0.01 | 0.41 | 0.70 | 0.91 |
| MCMC | $\tilde{\rho}$ | 0.00 | 0.41 | 0.70 | 0.89 |
| Two-step | Cor($\hat{\theta}_1, \hat{\theta}_2$) | 0.00 | 0.27 | 0.44 | 0.59 |

Table 5: Correlation estimates for 5 dimensions and 30 items

| Method | | $\rho$ 0.00 | 0.40 | 0.70 | 0.90 |
|---|---|---|---|---|---|
| Generated | Cor($\theta_1, \theta_2$) | -0.01 | 0.44 | 0.70 | 0.90 |
| MCMC | $\tilde{\rho}$ | -0.01 | 0.43 | 0.70 | 0.91 |
| Two-step | Cor($\hat{\theta}_1, \hat{\theta}_2$) | -0.01 | 0.37 | 0.60 | 0.79 |

Table 6: Correlation estimates for 5 dimensions and 50 items

| Method | | $\rho$ 0.00 | 0.40 | 0.70 | 0.90 |
|---|---|---|---|---|---|
| Generated | Cor($\theta_1, \theta_2$) | -0.02 | 0.42 | 0.68 | 0.90 |
| MCMC | $\tilde{\rho}$ | -0.02 | 0.43 | 0.68 | 0.89 |
| Two-step | Cor($\hat{\theta}_1, \hat{\theta}_2$) | -0.02 | 0.39 | 0.62 | 0.81 |

## 6.2 Estimates of ability

## 6.3 Correlation with true ability

Tables 7 through 12 list the correlations between the true ability and the estimated ability. When no correlation exists between abilities, the EAP-M and EAP-U estimates correlate equally well with the generating parameters. But as the correlation between abilities increases, the EAP-M estimates correlate more highly with the true ability whereas the EAP-U estimates are unaffected. As the number of items increases, the correlation between the true ability and the estimated ability for both methods also increases. Finally, increasing the number of abilities gives higher correlations for the EAP-M estimates but has has no impact on the EAP-U estimates.

Table 7: Correlations between the true and estimated abilities for 2 dimensions and 10 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.81 | 0.83 | 0.84 | 0.87 |
| EAP-U | 0.81 | 0.82 | 0.81 | 0.82 |

Table 8: Correlations between the true and estimated abilities for 2 dimensions and 30 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.92 | 0.93 | 0.93 | 0.94 |
| EAP-U | 0.94 | 0.93 | 0.92 | 0.92 |

Table 9: Correlations between the true and estimated abilities for 2 dimensions and 50 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.95 | 0.95 | 0.96 | 0.96 |
| EAP-U | 0.95 | 0.95 | 0.95 | 0.95 |

Table 10: Correlations between the true and estimated abilities for 5 dimensions and 10 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.80 | 0.83 | 0.86 | 0.91 |
| EAP-U | 0.80 | 0.81 | 0.79 | 0.80 |

Table 11: Correlations between the true and estimated abilities for 5 dimensions and 30 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.92 | 0.93 | 0.94 | 0.96 |
| EAP-U | 0.92 | 0.93 | 0.93 | 0.93 |

Table 12: Correlations between the true and estimated abilities for 5 dimensions and 50 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.95 | 0.95 | 0.96 | 0.97 |
| EAP-U | 0.95 | 0.95 | 0.95 | 0.95 |

## 6.4 Root mean squared error

The number of abilities, number of items, and the degree of correlation affect the RMSE of the EAP-M and EAP-U estimates in the same way that they affect the correlations between the true and estimated abilities. That is, (a) the two methods yield equivalent results when the abilities are not correlated; (b) a greater number of abilities or a greater degree of correlation between the abilities improves the EAP-M estimates but does not affect the EAP-U estimates; and (c) an increase in the number of items results in more precise estimates by both methods.

Table 13: RMSE of ability estimates for 2 dimensions and 10 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.59 | 0.57 | 0.54 | 0.49 |
| EAP-U | 0.59 | 0.58 | 0.58 | 0.57 |

Table 14: RMSE of ability estimates for 2 dimensions and 30 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.39 | 0.38 | 0.37 | 0.35 |
| EAP-U | 0.39 | 0.39 | 0.39 | 0.39 |

Table 15: RMSE of ability estimates for 2 dimensions and 50 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.31 | 0.30 | 0.29 | 0.27 |
| EAP-U | 0.31 | 0.30 | 0.30 | 0.31 |

Table 16: RMSE of ability estimates for 5 dimensions and 10 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.59 | 0.56 | 0.51 | 0.43 |
| EAP-U | 0.59 | 0.59 | 0.61 | 0.62 |

Table 17: RMSE of ability estimates for 5 dimensions and 30 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.38 | 0.38 | 0.34 | 0.29 |
| EAP-U | 0.38 | 0.39 | 0.38 | 0.37 |

Table 18: RMSE of ability estimates for 5 dimensions and 30 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.30 | 0.30 | 0.28 | 0.24 |
| EAP-U | 0.30 | 0.30 | 0.30 | 0.31 |

### 6.5 Efficiency

The posterior variance is approximately equal to the squared standard error of the estimate. The average posterior variance for each method is obtained by: (a) computing the variance of the last 8,000 draws for each examinee; (b) averaging the variances across the 1,000 examinees; and (c) averaging the variance again across the different abilities. Efficiency is defined in this paper as the average posterior variance of the EAP-U estimates over the average posterior variance of the EAP-M estimates. Thus, a ratio greater than 1.00 is interpreted as the EAP-M method being more efficient than the EAP-U method, and vice versa. In addition, the ratio also indicates the factor by which the test length needs to be increased for the EAP-U estimates to have the same precision as the EAP-M estimates obtained with the original test length.

When only two dimensions are concurrently considered, the efficiency of EAP-M method is not evident unless the abilities are very highly correlated (i.e., $\rho = 0.90$). Efficiency at this level ranges from 1.25 to 1.40. Depending on the length of the test, this is equivalent to adding 3 to 12 items to the test.

When five dimensions are simultaneously considered, the efficiency of the EAP-M method is evident for abilities that are reasonably highly correlated; efficiency ranges from 1.16 to 2.07. For some tests, the precision of the EAP-M estimates is equivalent to the precision of the EAP-M estimates obtained from tests that are twice as long. Depending on the original test length, this is equivalent to adding 4 to 27 item to test.

The increase in precision is less evident when long tests are used. This is consistent with the results discussed earlier which indicate that marginal improvement is slight when abilities are already well estimated. Although, efficiency may not be as high for long tests, the corresponding number of additional items turn out to be larger. Finally, for a fixed level of correlation between abilities that is greater than zero, the efficiency obtained from simultaneously using five dimensions are consistently higher compared to the efficiency obtained from using only two dimensions.

Table 19: Posterior variance of ability estimates for 2 dimensions and 10 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.34 | 0.33 | 0.31 | 0.25 |
| EAP-U | 0.34 | 0.34 | 0.34 | 0.34 |
| Efficiency | 1.00 | 1.03 | 1.10 | 1.34 |

Table 20: Posterior variance of ability estimates for 2 dimensions and 30 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.15 | 0.15 | 0.14 | 0.11 |
| EAP-U | 0.15 | 0.15 | 0.15 | 0.15 |
| Efficiency | 1.01 | 1.01 | 1.12 | 1.40 |

Table 21: Posterior variance of ability estimates for 2 dimensions and 50 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.09 | 0.09 | 0.09 | 0.07 |
| EAP-U | 0.09 | 0.09 | 0.09 | 0.09 |
| Efficiency | 0.99 | 1.02 | 1.07 | 1.25 |

Table 22: Posterior variance of ability estimates for 5 dimensions and 10 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.36 | 0.33 | 0.25 | 0.17 |
| EAP-U | 0.36 | 0.36 | 0.36 | 0.36 |
| Efficiency | 1.00 | 1.09 | 1.45 | 2.07 |

Table 23: Posterior variance of ability estimates for 5 dimensions and 30 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.14 | 0.14 | 0.12 | 0.08 |
| EAP-U | 0.14 | 0.14 | 0.14 | 0.14 |
| Efficiency | 1.00 | 1.05 | 1.23 | 1.86 |

Table 24: Posterior variance of ability estimates for 5 dimensions and 50 items

| | $\rho$ | | | |
|---|---|---|---|---|
| Method | 0.00 | 0.40 | 0.70 | 0.90 |
| EAP-M | 0.09 | 0.09 | 0.08 | 0.06 |
| EAP-U | 0.09 | 0.09 | 0.09 | 0.09 |
| Efficiency | 0.99 | 1.03 | 1.16 | 1.54 |

# 7 Discussion

The results show that the proposed multidimensional approach to simultaneous ability estimation gives more general outcomes. The method gives results similar to those of the unidimensional method when abilities are uncorrelated. When abilities are correlated, taking the correlation into account can lead to noticeable improvements in ability estimates, especially when there are multiple short tests and the underlying correlation is high. In addition, this hierarchical approach allows for direct estimation of the correlation between the abilities. This obviates the need for a two-step approach that leads to biased estimates (Little & Rubin, 1983; Mislevy, 1984), which in this case, are underestimates.

The multidimensional approach should be beneficial in many testing situations. The administration of multiple tests during one sitting is not uncommon, and as Johnson and Carlson (1994) reported, the different abilities measured by these tests are usually highly correlated. Although some of the improvement using this approach are relatively modest, it can be achieved without much additional cost, (i.e., only the estimation process was changed in scoring the same data sets). In a practical sense, use of this method means that, given a fixed number of items, ability estimates can be made more precise, or given a desired level of precision, the number of items can be reduced without loss of accuracy.

This method can also be applied to a single test composed of several subtests. Currently, there is a great interest in using test results for diagnostic purposes (i.e., determining the students' strong and weak points). However, in many instances this objective is not realized because separate scores cannot be reported due to insufficient reliability of the subtests. Hence, the continued reliance on a single, more global composite score (Wainer et al, 2001). This could be a promising application given the nature of the composite tests (i.e., multiple short sections that are highly correlated).

It should be noted that the assumption of simple structure does not limit the usefulness of the proposed method. On the contrary, the assumption makes the application of the method more straightforward in that it can be applied without changes in the item response models to existing tests that have already been calibrated.

Future research might take a variety of directions. First, although the results show that the proposed method works for simulated data sets, it is important to verify that it works for real-world data as well. Second, the present paper uses item parameters with known values. The approach can be broadened to include item parameter estimation such as was done by Patz and Junker (1999 a,b) in the unidimensional IRT case. Finally, the proposed method can be tried with other item response models such as the generalized graded unfolding model of Roberts, Donoghue and Laughlin (2000) models, and other testing contexts.

# References


[1] Casella, G., & George, E. I. (1995). Explaining the Gibbs sampler. *The American Statistician, 46*, 167-174.

[2] Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician, 49*, 327-335.

[3] Gamerman, D. (1996). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference.* London: Chapman & Hall.

[4] Gelman, A., Carlin, J. B., Stern, H., & Rubin, D. B. (1996). *Bayesian data analysis.* London: Chapman & Hall.

[5] Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). Introducing Markov chain Monte Carlo. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), In *Markov chain Monte Carlo in practice* (pp 1-17). London: Chapman & Hall.

[6] Johnson. E. G., & Carlson, J. (1994). *The NAEP 1992 Technical Report* (Report No. 23-TR-20). Washington, DC: National Center for Education Statistics.

[7] Little, R. J. A., & Rubin, D. B. (1983). On jointly estimating parameters and missing data by maximizing the complete-data likelihood. *The American Statistician, 37*, 218-220.

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

[9] Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika, 49*, 359-381.

[10] Patz, R. J., and Junker, B. W. (1999a). A straightforward approach to Markov chain Monte Carlo methods for item response theory. *Journal of Educational and Behavioral Statistics, 24*, 146-178.

[11] Patz, R. J., and Junker, B. W. (1999b). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics, 24*, 342-366.

[12] Reckase, M. D. (1996). A linear logistic multidimensional model. In W. J. van der Linder & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp 271-286). New York: Springer-Verlag.

[13] Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general model for unfolding unidimensional polytomous responses using item response theory. *Applied Psychological Measurement, 24*, 3-32.

[14] Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics, 22*, 1701-1762.

[15] Wainer, H., Vevea, J. L., Camacho, F., Reeve III, B. B., Rosa, K., Nelson, L., Swygert, K. A., & Thissen, D. (2001). Augmented scores - "Borrowing strength" to compute score based on small numbers of items. In D. Thissen, & H. Wainer (Eds.). *Test scoring* (pp. 343-388). Mahwah, NJ: Erlbaum.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

# REPRODUCTION RELEASE

TM033825

(Specific Document)

## I. DOCUMENT IDENTIFICATION:

| | |
|---|---|
| Title: A Multidimensional Item Response Theory Approach to Simultaneous Ability Estimation | |
| Author(s): Jimmy de la Torre and Richard J. Patz | |
| Corporate Source: University of Illinois at Urbana-Champaign | Publication Date: April 2002 |

## II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

| The sample sticker shown below will be affixed to all Level 1 documents | The sample sticker shown below will be affixed to all Level 2A documents | The sample sticker shown below will be affixed to all Level 2B documents |
|---|---|---|
| PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1 | PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A | PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B |
| Level 1 ☑ | Level 2A ☐ | Level 2B ☐ |
| Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy. | Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only | Check here for Level 2B release, permitting reproduction and dissemination in microfiche only |

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

*I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.*

**Sign here, → please**

| | | |
|---|---|---|
| Signature: | Printed Name/Position/Title: JIMMY DE LA TORRE / GRADUATE STUDENT | |
| Organization/Address: 603 E. Daniel St. Champaign IL 61820 | Telephone: (217) 265-0384 | FAX: |
| | E-Mail Address: jdelator@s-psych.uiuc.edu | Date: 4/2/02 |

(over)

## III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

| |
|---|
| Publisher/Distributor: |
| Address: |
| Price: |

## IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

| |
|---|
| Name: |
| Address: |

## V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

**ERIC Processing and Reference Facility**
**4483-A Forbes Boulevard**
**Lanham, Maryland 20706**

**Telephone: 301-552-4200**
**Toll Free: 800-799-3742**
**FAX: 301-552-4700**
**e-mail: ericfac@inet.ed.gov**
**WWW: http://ericfac.piccard.csc.com**

EFF-088 (Rev. 2/2000)