Comparison of Adaptive Bayesian Estimation and Weighted Bayesian Estimation in Multidimensional Computerized Adaptive Testing

Po Hsi Chen

Department of Educational Psychology and Counseling, National Taiwan Normal University

Presented at the Multidimensional CAT Paper Session, June 3, 2009

Abstract

The goal of this study was to compare the robustness of two modified Bayesian estimation methods with the traditional Bayesian estimation in multidimensional computerized adaptive testing (MCAT). The independent variable was the ability (θ) estimation methods: adaptive Bayesian maximum a posteriori (AMAP) estimation, weighted Bayesian maximum a posteriori (WMAP) estimation, and traditional Bayesian maximum a posteriori estimation (MAP). The dependent variables were conditional bias and root mean square error (RMSE). Ten to sixty items were used in two-dimensional computerized adaptive testing (CAT) with the three θ estimation methods in order to investigate the precision of θ estimation. The robustness of the three methods to the prior distribution were also been compared. Results indicated that the AMAP and WMAP can effectively reduce the regression bias of MAP estimation. The RMSE using AMAP and WMAP were larger than using the traditional MAP estimation when the θ combinations were correspondent with the prior distribution; however, the RMSE of WMAP become close to that of MAP when the test length increased. When the θ combinations were not correspondent with the prior distribution, the RMSE using WMAP was lower than using the MAP and AMAP. WMAP was much more robust to the prior distribution than traditional MAP.

Acknowledgment

This research was supported by the National Science Council in Taiwan (NSC 95-2413-H-003-044). The author acknowledge the funding of Taiwan NSC. Presentation of this paper at the 2009 Conference on Computerized Adaptive Testing was supported in part with funds from GMAC®.

Citation

Chen, P. H. (2009). Comparison of adaptive Bayesian estimation and weighted Bayesian estimation in multidimensional computerized adaptive testing. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing.* Retrieved [date] from www.psych.umn.edu/psylabs/CATCentral/

Author Contact

**Department of Educational Psychology and Counseling, National Taiwan Normal University, No. 162, Hoping E. Rd., Sec. 1, Taipei 106, Taiwan.
Email: chenph@ntnu.edu.tw**

Comparison of Adaptive Bayesian Estimation and Weighted Bayesian Estimation in Multidimensional Computerized Adaptive Testing

Introduction

The precision of ability (θ) estimation in item response theory (IRT) and computerized adaptive testing (CAT) is related to the estimation method. Research in unidimensional IRT and CAT has indicated that maximum likelihood (ML) estimation resulted in higher standard errors and root mean square error (RMSE) than the Bayesian estimation methods (Bock & Mislevy, 1982; Weiss & Mcbride, 1984). Bayesian maximum a posteriori (MAP) estimation and Bayesian expected a posteriori (EAP) estimation resulted in lower standard error and RMSE but higher regression bias. In other words, the θ estimates of the examinees with extreme θ levels were biased toward the mean of the prior distribution.

The procedures of θ estimation methods of multidimensional CAT (MCAT) are more complex than those of unidimensional CAT (UCAT). Segall (1996) demonstrated that when the UCAT is extended to MCAT, the goal of ML estimation is to find the θ vector that maximizes the likelihood function,

$$L(\mathbf{u}_n \mid \boldsymbol{\theta}_s) = \prod_{i=1}^{n} P(\boldsymbol{\theta}_s)^{u_i} Q(\boldsymbol{\theta}_s)^{1-u_i} , \tag{1}$$

where $L(\mathbf{u}_n \mid \boldsymbol{\theta}_s)$ is the likelihood function when the examinee's temporary θ vector is $\boldsymbol{\theta}_s$, $P(\boldsymbol{\theta}_s)$ is the probability of endorsing an item, and $Q(\boldsymbol{\theta}_s) = 1 - P(\boldsymbol{\theta}_s)$. Newton-Raphson iterations can be used to find the $\boldsymbol{\theta}_s$ vector quickly. The Newton-Raphson iteration method in MCAT is modified as (Segall, 1996)

$$\boldsymbol{\theta}^{(n)} = \boldsymbol{\theta}^{(n-1)} - [\frac{\partial^2 \ln L}{\partial \boldsymbol{\theta}^2}]^{-1} \frac{\partial \ln L}{\partial \boldsymbol{\theta}} , \tag{2}$$

where $\boldsymbol{\theta}^{(n)}$ is the examinee's temporary θ vector at the n^{th} iteration, and $\boldsymbol{\theta}^{(n-1)}$ is the examinee's temporary θ vector at the $n-1^{\text{th}}$ iteration. The first and the second derivatives of the likelihood function are

$$\frac{\partial \ln L}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \ln L}{\partial \theta_1} \\ \frac{\partial \ln L}{\partial \theta_2} \\ \vdots \\ \frac{\partial \ln L}{\partial \theta_d} \end{bmatrix}, \tag{3}$$

$$\frac{\partial^2 \ln L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial^2 \ln L}{\partial \theta_1^{\ 2}} & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_d} \\ \frac{\partial^2 \ln L}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_2^{\ 2}} & \cdots & \frac{\partial^2 \ln L}{\partial \theta_2 \partial \theta_d} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 \ln L}{\partial \theta_d \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_d \partial \theta_2} & \cdots & \frac{\partial^2 \ln L}{\partial \theta_d^{\ 2}} \end{bmatrix}, \tag{4}$$

where L was shown in Equation 1 , $\frac{\partial \ln L}{\partial \theta_d}$ is the first partial derivative of L by θ_d, and $\frac{\partial^2 \ln L}{\partial \theta_d \partial \theta_s}$ is the second partial derivative of L by θ_d and θ_s.

The Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM)

The MRCMLM (Adams, Wilson, & Wang, 1997) is a non-compensatory multidimensional Rasch model. It calibrates item and person parameters jointly and yields a direct estimate of the person population variance-covariance matrix. Under the MRCMLM, the probability of a response in category k of item i for a person with vector $\boldsymbol{\theta}$ across d dimensions is expressed as

$$f_{ik}(\boldsymbol{\theta}) \equiv f\left(X_{ik} = 1; \boldsymbol{\xi} \mid \boldsymbol{\theta}\right) = \frac{\exp\left(\mathbf{b}_{ik}^{'}\boldsymbol{\theta} + \mathbf{a}_{ik}^{'}\boldsymbol{\xi}\right)}{\sum_{u=1}^{K_i} \exp\left(\mathbf{b}_{iu}^{'}\boldsymbol{\theta} + \mathbf{a}_{iu}^{'}\boldsymbol{\xi}\right)}, \tag{5}$$

where

$$X_{ik} = \begin{cases} 1 & \text{if response to item } i \text{ is in category } k, \\ 0 & \text{otherwise.} \end{cases}$$

$\boldsymbol{\xi}$ is a vector of p parameters that describes the items,

$\boldsymbol{\theta}$ is vector of d-dimensional latent traits that describes a particular person,

$\mathbf{a}_{ik}$, for category k of item i, each of length p, is the linear combination coefficients of $\boldsymbol{\xi}$,

$\mathbf{b}_{ik}$ is a vector of scores for category k of item i, across d dimensions,

and K_i is the number of response categories in item i.

The reason for using the model in this research is that it is a general model of many existing dichotomous and polytomous unidimensional Rasch models, such as the simple logistic model (Rasch, 1960), the linear logistic latent trait model (Fischer, 1973), the rating scale model (Andrich, 1978), the partial credit model (Masters, 1982), the facet model (Linacre, 1989), the ordered partitioned model (Wilson, 1992), and the linear partial credit model (Fischer & Ponocny, 1994). More importantly, the formulation allows the specification of a range of multidimensional models to be used in MCAT, such as the multidimensional rating scale model and the partial credit model. Being a member of the Rasch family, the MRCML shares the same desirable measurement properties, such as specific objectivity and interval scales. The ACER ConQuest software (Wu, Adams, & Wilson, 1998) for the MRCML uses marginal maximum likelihood estimation with Bock and Aitkin's (1981) formulation of the EM algorithm (Dempster, Laird, & Rubin, 1977). All the items parameters, $\boldsymbol{\xi}$, and person population parameters $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$, can be simultaneously estimated.

The θ Estimation Methods of MCAT Using MRCMLM

When applying the MRCMLM to the MCAT procedure (Wang & Chen, 2004), the elements of the 1st and 2nd derivatives of the likelihood function for the ML estimation are expressed as

$$\frac{\partial \ln L(\boldsymbol{\theta})}{\partial \theta_d} = \frac{\left[b_{ik} f_{ik}(\boldsymbol{\theta}) - f_{ik}(\boldsymbol{\theta}) \sum_{k=1}^{K} b_{ik} f_{ik}(\boldsymbol{\theta}) \right]}{f_{ik}(\boldsymbol{\theta})}, \tag{6}$$

$$\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial \theta_d \partial \theta_s} = \sum_{k=1}^{K} [b_{ik} b_{ik}' f_{ik}(\boldsymbol{\theta}) - \sum_{k=1}^{K} b_{ik} f_{ik}(\boldsymbol{\theta}_d) \sum_{k=1}^{K} b_{ik} f_{ik}(\boldsymbol{\theta}_s)], \tag{7}$$

where $L(\boldsymbol{\theta})$ denotes the likelihood function when the examinee's temporary ability vector is $\boldsymbol{\theta}$, $f_{ik}(\boldsymbol{\theta})$ is shown in Equation 5. b_{ik} is the element of the $\mathbf{b}_{ik}$ vector in Equation 5. When these 1st and 2nd derivatives are used in the Newton-Raphson iteration procedure, the temporary vector $\boldsymbol{\theta}$ that maximizes the likelihood function can be estimated. The item selection procedure, which was suggested by Segall(1996), can therefore be used to decide the next item to be administered.

When using Bayesian maximum a posteriori (MAP) estimation, the prior distribution of the population should be taken into consideration to form the posterior density function, which is expressed as

$$f(\boldsymbol{\theta} \mid \mathbf{X}) = \frac{L(\mathbf{X} \mid \boldsymbol{\theta}) g(\boldsymbol{\theta})}{f(\mathbf{X})}, \tag{8}$$

$L(\mathbf{X}|\boldsymbol{\theta})$ denotes the likelihood of observing response vector **X** given $\boldsymbol{\theta}$, $g(\boldsymbol{\theta})$ denotes the prior distribution of $\boldsymbol{\theta}$, and $f(\mathbf{X})$ denotes the marginal probability of **X**. To maximize the posterior density function, the same Newton-Raphson iteration procedure can be used. The elements of the 1st and 2nd derivatives of the posterior density function for MAP estimation are expressed as (Wang & Chen, 2004):

$$\frac{\partial \ln f(\boldsymbol{\theta} \mid \mathbf{X})}{\partial \theta_d} = \frac{\left[b_{ik} f_{ik}(\boldsymbol{\theta}) - f_{ik}(\boldsymbol{\theta}) \sum_{k=1}^{K} b_{ik} f_{ik}(\boldsymbol{\theta}) \right]}{f_{ik}(\boldsymbol{\theta})} - \left[\frac{\partial (\boldsymbol{\theta} - \boldsymbol{\mu})'}{\partial \theta_d} \right] \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}), \tag{9}$$

$$\frac{\partial^2 \ln f(\boldsymbol{\theta} \mid \mathbf{X})}{\partial \theta_d \partial \theta_s} = \sum_{k=1}^{K} [b_{ik} b_{ik}' f_{ik}(\boldsymbol{\theta}) - \sum_{k=1}^{K} b_{ik} f_{ik}(\boldsymbol{\theta}_d) \sum_{k=1}^{K} b_{ik} f_{ik}(\boldsymbol{\theta}_s)] - \phi_{ds}^{-1}, \tag{10}$$

where $\boldsymbol{\mu}$ denotes the mean vector of the prior distribution, $\boldsymbol{\Sigma}^{-1}$ denotes the inverse of the variance-covariance matrix, and ϕ_{ds}^{-1} denotes the (d, s)th element of $\boldsymbol{\Sigma}^{-1}$.

When using Bayesian expected a posteriori (EAP) estimation, no iteration procedure is needed, although the expected vector $\boldsymbol{\theta}_{EAP}$ of the posterior density function must be computed. The $\boldsymbol{\theta}_{EAP}$ is estimated by

$$\boldsymbol{\theta}_{EAP} = \sum_{q=1}^{k_q} \boldsymbol{\theta}_q f(\boldsymbol{\theta}_q \mid \mathbf{X}) = \sum_{q=1}^{k_q} \boldsymbol{\theta}_q \frac{L(\mathbf{X} \mid \boldsymbol{\theta}_q) g(\boldsymbol{\theta}_q)}{\sum_{q=1}^{k_q} [L(\mathbf{X} \mid \boldsymbol{\theta}_q) g(\boldsymbol{\theta}_q)]}, \tag{11}$$

where k_q denotes the quadrature points of each θ dimension when computing the expected value of the posterior density function, $L(\mathbf{X} \mid \boldsymbol{\theta}_q)$ denotes the value of the likelihood function at each θ combination point, $\boldsymbol{\theta}_q$, and $g(\boldsymbol{\theta}_q)$ denotes the value of the prior distribution of $\boldsymbol{\theta}$. In EAP estimation, the numbers of the θ combinations of $\boldsymbol{\theta}_q$ to be averaged is k_q to the power of the number of dimensions. For example, in the condition of the four-dimensional test with $k_q = 30$ in each dimension, the total number of combinations of $\boldsymbol{\theta}_q$ to be averaged are 810, 000(= 30^4). This means that in higher-dimensional CAT, EAP estimation is infeasible because it takes too much computation time. MAP estimates of $\boldsymbol{\theta}_{.}$ require far less computation time than EAP estimation for problems of higher dimensionality.

Chen (2006) compared ML, MAP, and EAP estimation in MCAT using MRCMLM and found that the regression bias of MAP estimation was more serious in MCAT than in UCAT, especially when test length was short and the correlations among the dimensions were high. When using MAP and EAP estimation, the RMSE of the θ estimates were lower than for ML estimation. When the number of θ dimensions was increased to four in MCAT, EAP estimation was infeasible because it took too much time. For example, about 15 seconds were required to estimate the temporary $\boldsymbol{\theta}_{EAP}$ and to select the next item with 30 quadrature points for each θ dimension. ML estimation is also infeasible when the examinee obtained a perfect or zero score on one of the dimensions.

By using the information of the prior distribution, MAP estimation can provide lower estimated error of θ. However, the use of prior information would bias the θ estimators to the centroid of the prior distribution in MCAT. Hence, it is important to develop a procedure to reduce the regression bias without losing the precision of θ estimation of the MAP estimation in MCAT. Two procedures have been suggested to modify Bayesian estimation in MCAT: adaptive Bayesian estimation (Raiche, Blais, & Magis, 2007) weighted Bayesian estimation (Chen, 2007). The goal of this research was to compare the precision of θ estimation of these two modified Bayesian estimation methods with traditional Bayesian estimation in MCAT.

Adaptive Bayesian Estimation

In adaptive Bayesian estimation (AMAP), the prior distribution of $\boldsymbol{\theta}$ is changed after every item is administered in MCAT. Hence, the adaptive MAP estimator after k items are administered is modified as

$$\hat{\boldsymbol{\theta}}_k = MAX_{\boldsymbol{\theta}}\{L(U \mid \hat{\boldsymbol{\theta}})P(\hat{\boldsymbol{\theta}})_{k-1}\}, \tag{12}$$

where $L(U \mid \hat{\boldsymbol{\theta}})$ is the likelihood function after k items, and $P(\hat{\boldsymbol{\theta}})_{k-1}$ is the modified prior distribution after $k-1$ items. To simplify the procedure in order to save the computation timesin MCAT, only the prior mean vector was modified after every item was administered; the variance-covariance matrix of the prior distribution was fixed in the research of Raiche, Blais, and Magis (2007).

Weighted Bayesian Estimation

In weighted Bayesian estimation (WMAP), a weighted value is added to the prior distribution to diminish the influences of the prior information. The posterior density function of WMAP is

$$f(\boldsymbol{\theta} \mid \mathbf{X}) = \frac{L(\mathbf{X} \mid \boldsymbol{\theta})w(\hat{\boldsymbol{\theta}})g(\boldsymbol{\theta})}{f(\mathbf{X})}. \tag{13}$$

In MCAT, the weighted value of examinee $\boldsymbol{\theta}$ after k items is

$$w(\hat{\boldsymbol{\theta}})_k = \frac{1}{1 + M(\hat{\boldsymbol{\theta}})_{k-1}}, \tag{14}$$

$$M(\hat{\boldsymbol{\theta}})_{k-1} = \sqrt{(\hat{\boldsymbol{\theta}}_{k-1} - \boldsymbol{\mu})' \Phi^{-1} (\hat{\boldsymbol{\theta}}_{k-1} - \boldsymbol{\mu})}, \tag{15}$$

where $M(\hat{\boldsymbol{\theta}})_{k-1}$ is the Mahalanobis distance (Mahalanobis, 1936) between the temporary θ estimators and the centroid of the prior distribution after *k*-1 items, and $M(\hat{\boldsymbol{\theta}})_0$ is 0. The denominator in Equation 14 is increased by 1 in order to make the weighting value smaller than 1. The Mahalanobis distance is used in the multivariate analysis to detect outliers (Johnson, 1998), which takes the variance and covariance between dimensions into consideration. When the dimensions decrease to1, it becomes a *Z* score. Therefore, WMAP can apply not only in MCAT but also in UCAT. During the θ estimation process in MCAT, if the temporary θ vector is close to the prior mean vector, the weighted value is close to 1. If the temporary θ vector is distant from the prior mean vector, the weighted value is close to 0. However, if the weighted value is too small, WMAP estimation would become ML estimation and cannot apply to the examinees with a perfect score or zero score. Hence, the minimum weighted value is suggested.

WMAP is also robust to an unsuitable choice of the prior distribution. When the θ combination is correspondent with the prior distribution, the weighted values will be close to 1, and the prior information will be used completely just like in traditional MAP estimation. When the θ combination is not correspondent with the prior distribution, the weighted value is close to 0, and the prior information will become irrelevant to θ estimation. Since the θ estimator is changeable during the estimation process of CAT, the weighted value will be renewed after every item is administered. The *dynamic* change of the weighted value makes the WMAP more “adaptive” to all the examinees than traditional MAP estimation, regardless of whether the prior distribution is appropriate.

The goal of this research is to compare the precision of θ estimation using MAP, adaptive MAP (AMAP), and WMAP in MCAT, and to investigate the robust of the three MAP methods to the unsuitable choice of the prior distribution in MCAT.

Method

Research Design

MRCMLM was used to generate the responses. The independent variables were: (1) θ estimation methods—AMAP, WMAP, and MAP, and (2) the target MCAT test lengths—10, 20, 30, 40, 50, and 60 items. The dependent variables were the conditional bias and root mean square error (RMSE) of the θ estimates.

Data Generation

Two-dimensional θs were used in this research. The θ values were set as –3, –2, –1, 0, 1, 2, and 3 logits for each dimension. There were 49 (7 × 7) θ combinations, with 100 simulees in each combination. 400 difficulty parameters for dichotomous items were generated from the uniform distribution [U (–4.0, ~4.0)]. Half of the items were set to measure Dimension 1, and the others were set to measure Dimension 2. The response data were generated by the MRCMLM with the above θ combinations and item difficulties.

MCAT Procedures

MCATs of 10, 20, 30, 40, 50, and 60 items of MCAT were implemented using AMAP, WMAP, and MAP estimation. The correlation between θs was set at .8 in order to investigate their robustness to an unsuitable choice of the prior distribution. The mean vector of the prior distribution was set as (0, 0) and the variance-covariance matrix was specified as

1.00 0.80

0.80 1.00

By using this variance-covariance matrix, the prior distribution was suitable to only some θ combinations but was not suitable to the other θ combinations, especially to the outliers. Hence, the robust of the three MAP estimation methods to an unsuitable choice of the prior distribution could be evaluated .

Data Analysis

The conditional bias and root mean square error (RMSE) of the θ combination (*d*, *s*) was computed using the following:

$$bias(\hat{\theta}_{(d,s)}) = \frac{\sum_{k=1}^{100}(\hat{\theta}_{(d,s)k} - \theta_{(d,s)0})}{100} \tag{16}$$

$$RMSE(\hat{\theta}_{(d,s)}) = \sqrt{\frac{\sum_{k=1}^{100}(\hat{\theta}_{(d,s)k} - \theta_{(d,s)0})^2}{100}} \tag{17}$$

where $\theta_{(d,s)0}$ denotes the true value of the θ combination (*d*, *s*), and $\hat{\theta}_{(d,s)k}$ denotes the estimated value. The results of all θ combinations were classified into two groups. If the Mahalanobis distance between $\theta_{(d,s)0}$ and the centroid of the prior distribution was less than 3, the θ combination was classified as correspondent with the prior distribution. If the Mahalanobis distance between $\theta_{(d,s)0}$ and the centroid of the prior distribution was equal to or larger than 3, the θ combination was classified as not correspondent with the prior distribution. The result of the later group was to show the effect when the prior distribution was not suitable to the θ combination. Hence, the robustness of the estimation methods to the unsuitable choice of the prior distribution could be demonstrated.

Results

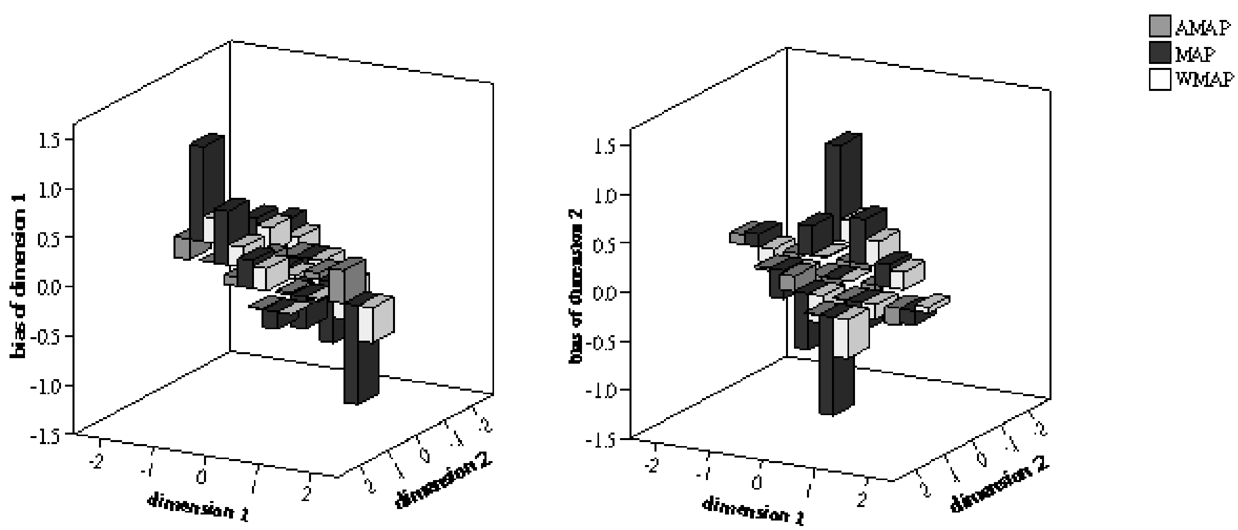
Bias

Figure 1 shows the conditional bias of 10, 30, and 50 item MCATs using AMAP, MAP, and WMAP estimation when the θ combinations were correspondent with the prior distribution. The results of 20, 40, and 60 items MCAT were similar to these results, so they are not shown here. When the MCAT test length was 10 items (Figure 1a), traditional MAP (dark bar) estimation resulted in regression bias—namely, the further the θ combination was away from the centroid of the prior distribution, the more bias there was to the centroid. For example, when the θ combination was (2, 1), the bias of dimension 1 and 2 were –1.01 and –.15 (the dark bars at the lower right area) respectively. When the θ combination was (–2, –1), the bias of Dimensions 1 and 2 were .96 and .13, respectively (the dark bars at the upper left). AMAP (gray bar) and WMAP (white bar) reduced the regression bias of the MAP and with lower bias than the MAP at most θ combinations.

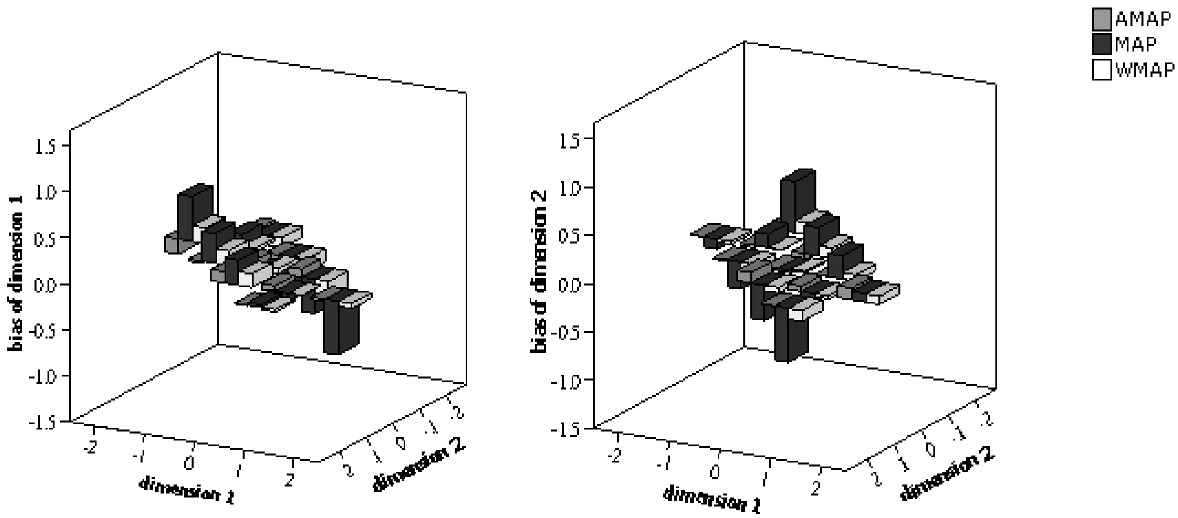
When the test length increased to 30 (Figure 1b) and 50 items (Figure 1c), the regression bias of all three MAP estimations gradually decrease. However, traditional MAP estimation still results in regression bias obviously. The AMAP and WMAP estimations show only little bias. These results indicated that the two modified MAP estimations can effectively reduce the regression bias of the traditional MAP estimation when the prior distribution is suitable to the examinees.

Figure 1. The Conditional Bias of MCAT When the θ Combinations Were Correspondent With the Prior Distribution

a. 10 Items



b. 30 Items



c. 50 Items

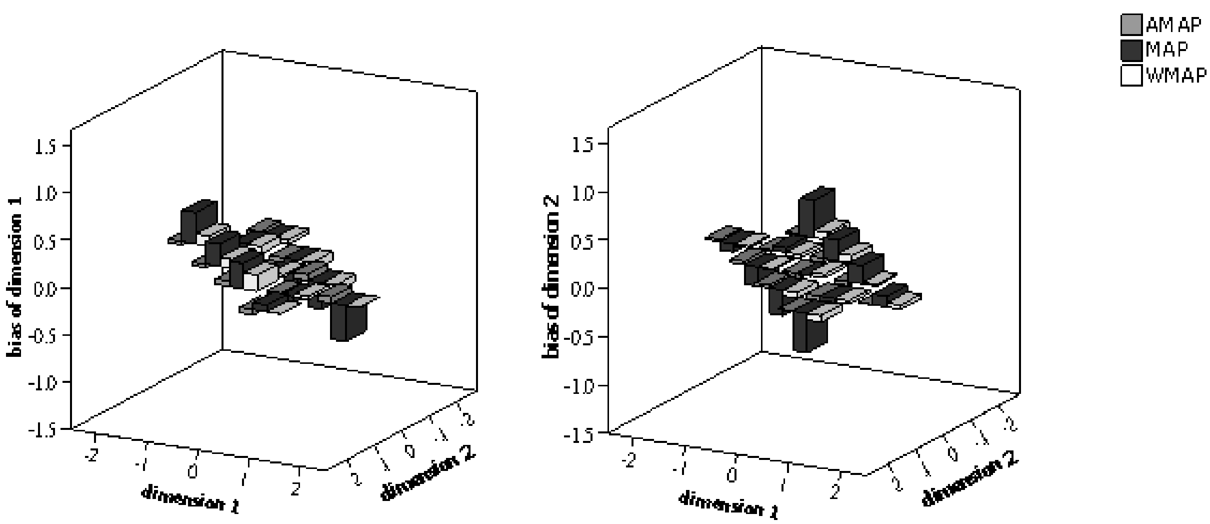
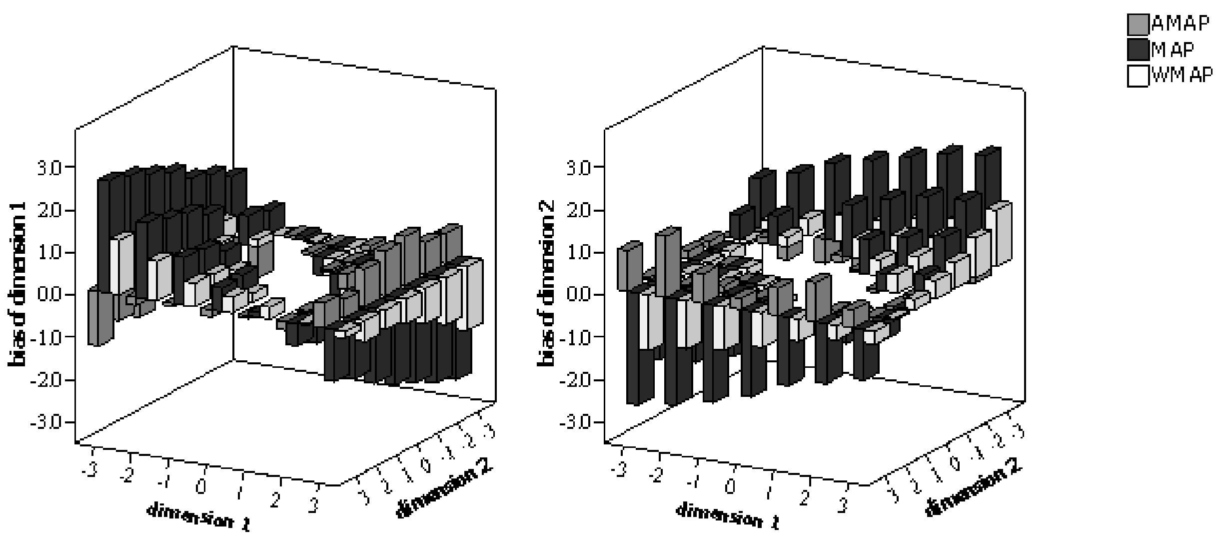


Figure 2 shows the conditional bias of 10, 30, and 50 items MCATs using AMAP, MAP, and WMAP when the θ combinations were not correspondent with the prior distribution. When the test length was 10 items (Figure 2a), the regression bias of traditional MAP estimation was much serious than that in Figure 1. In other words, if the prior distribution was not suitable to the θ combinations, using traditional MAP estimation will result in serious regression bias in MCAT. This agrees with the results of Chen (2006). However, the bias of AMAP and WMAP estimation were much lower than MAP, especially when the θ combinations are far from the centroid of the prior distribution. This result indicates that AMAP and WMAP estimations were more robust to the unsuitable choice of the prior distribution.

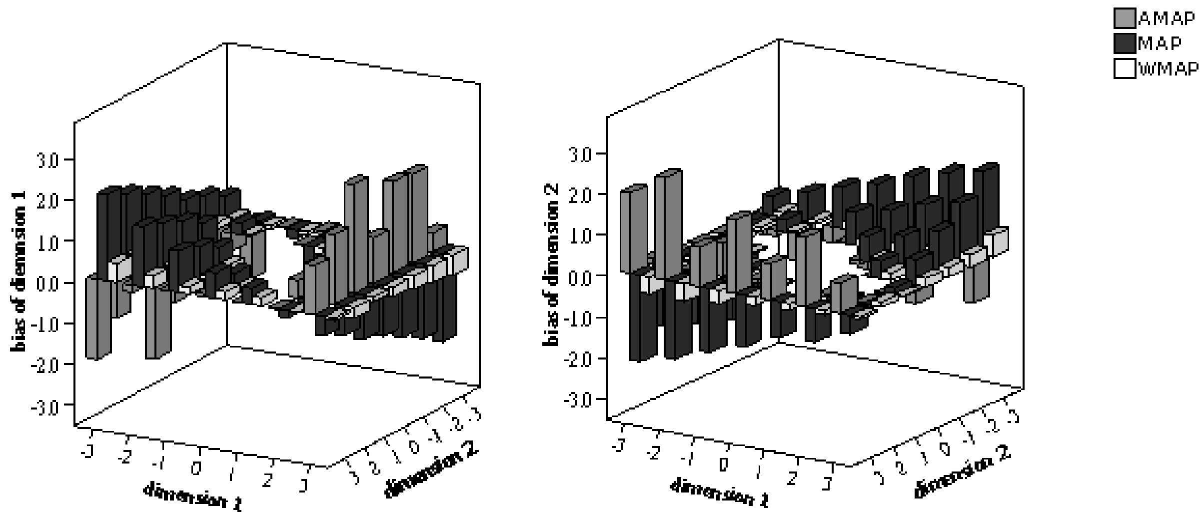
When the test length increased to 30 and 50 items (Figure 2a and 2b), the regression bias of all three MAP estimations gradually decreased. WMAP and AMAP still showed less bias than MAP estimation. However, at the extreme θ levels, AMAP was biased to the opposite side away from the centroid of the prior distribution. For example, when the θ values were 3.0 or –3.0, the bias (the gray bars) were high and in the reverse direction as the test length increased. The result indicates that AMAP estimation will over adjust the estimator and result in bias in the opposite direction at the extreme θ levels.

Figure 2. The Conditional Bias of MCAT When the θ Combinations Were Not Correspondent With the Prior Distribution

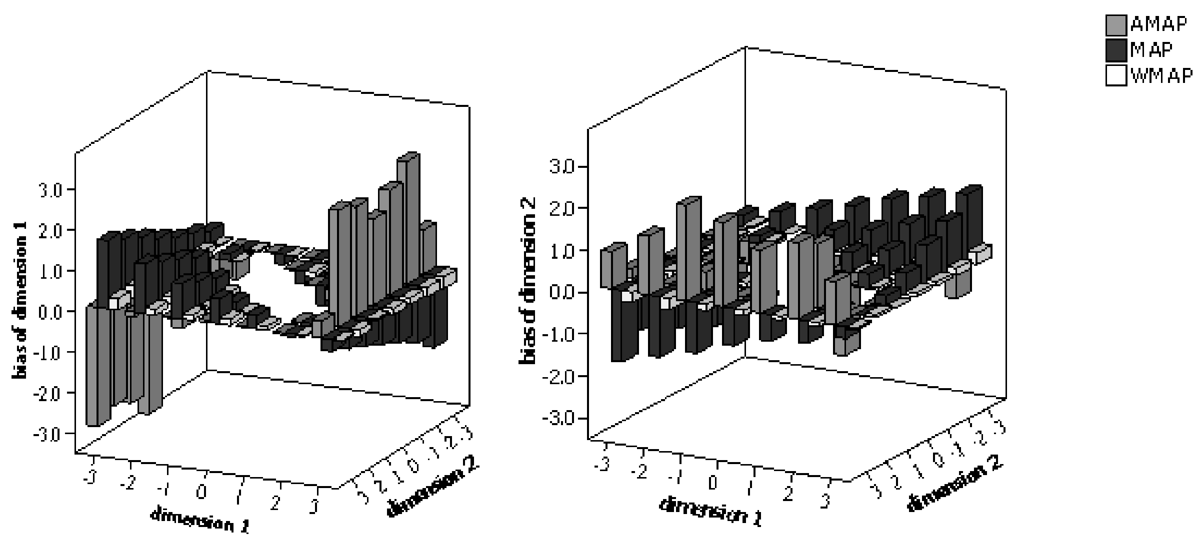
a. 10 Items



b. 30 Items



c. 50 Items



Summary. When the θ combinations were correspondent with the prior distribution, AMAP and WMAP showed less regression bias than MAP. The regression bias of all three MAP estimators decreased when the test length increased. When the θ combinations were not correspondent with the prior distribution, the regression bias of MAP was serious. AMAP and WMAP reduced the regression bias effectively and were robust to the unsuitable choice of the prior distribution. AMAP over-adjusted the θ estimates to the opposite side for the extreme θ levels.

RMSE

Table 1 shows the average RMSE for 10, 30, and 50 items using AMAP, MAP, and WMAP estimation. When the θ combinations were correspondent with the prior distribution, the average RMSE of MAP was lower than that of AMAP and WMAP. When the test length increased to 30 and 50 items, the average RMSE of the three MAP estimators decreased, and the average RMSE of WMAP and AMAP were close to that of MAP. When the θ combinations were not correspondent with the prior distribution, the average RMSE of WMAP was lower than MAP and AMAP. When test length increased, the average RMSE of MAP and WMAP decreased; however, the average RMSE of AMAP was very high. These results indicate that WMAP estimation was more robust to unsuitable choice of the prior distribution than the MAP and AMAP.

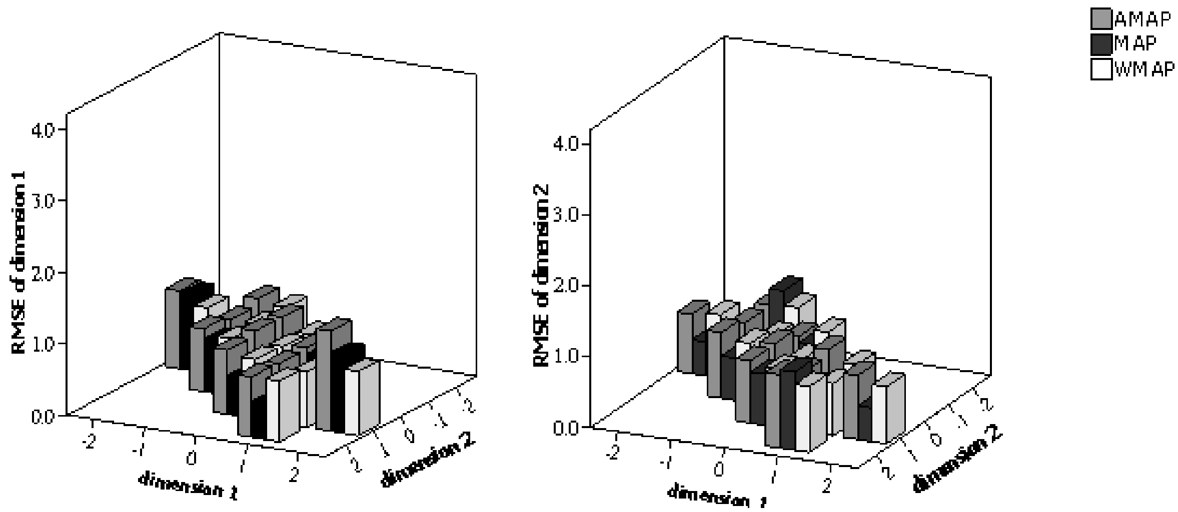
Table 1. The Average RMSE of MCAT Using Three MAP Estimators

		Dimension 1			Dimension 2		
θ Combination	Method	10 Items	30 items	50 items	10 items	30 items	50 items
Correspondent with the prior distribution	AMAP	0.94	0.59	0.43	0.88	0.54	0.42
	MAP	0.66	0.46	0.37	0.65	0.46	0.37
	WMAP	0.83	0.52	0.40	0.82	0.52	0.39
Not correspondent with the prior distribution	AMAP	1.24	1.42	1.55	1.15	1.33	1.47
	MAP	1.45	1.04	0.83	1.43	1.03	0.82
	WMAP	1.09	0.59	0.43	1.08	0.57	0.42

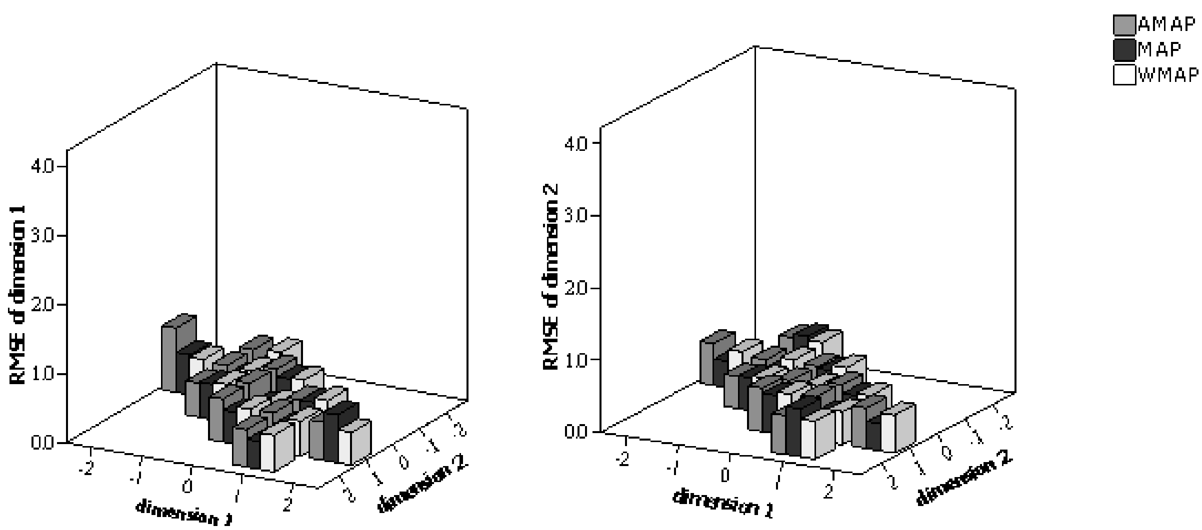
Figure 3 displays the conditional RMSE of 10, 30, and 50 items MCATs using AMAP, MAP, and WMAP when the θ combinations were correspondent with the prior distribution. When the test length was 10 items (Figure 3a), the RMSE of MAP (dark bars) wwa lower than that of WMAP (white bars) and AMAP (gray bars). When the test length increased, the RMSE of all three MAP estimators decreased, and the RMSE of AMAP and WMAP were very close to the results of MAP.

Figure 3. The Conditional RMSE When the θ Combinations Were Correspondent With the Prior Distribution

a. 10 items



a. 30 Items



b. 50 Items

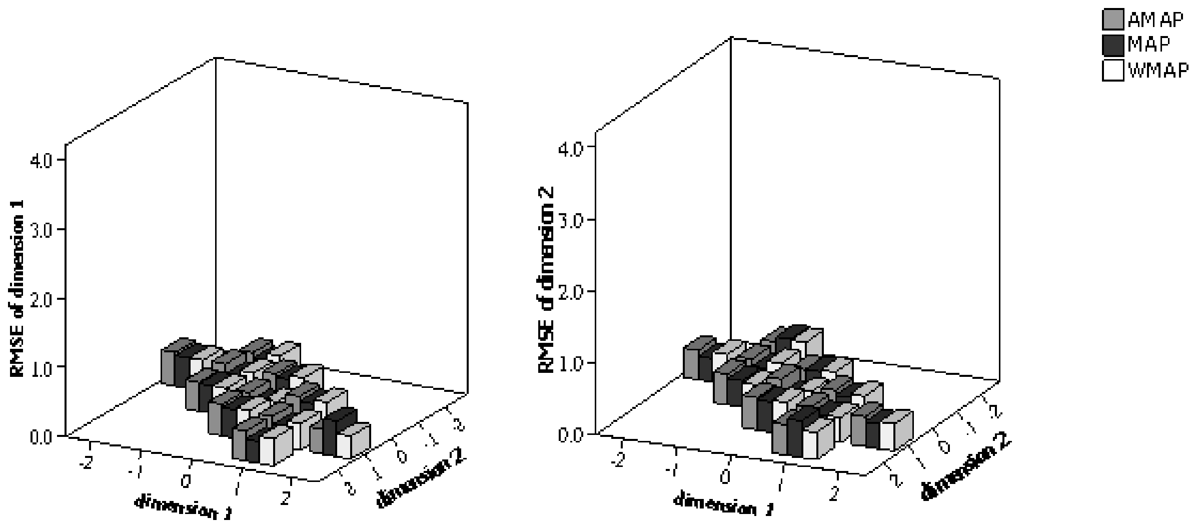
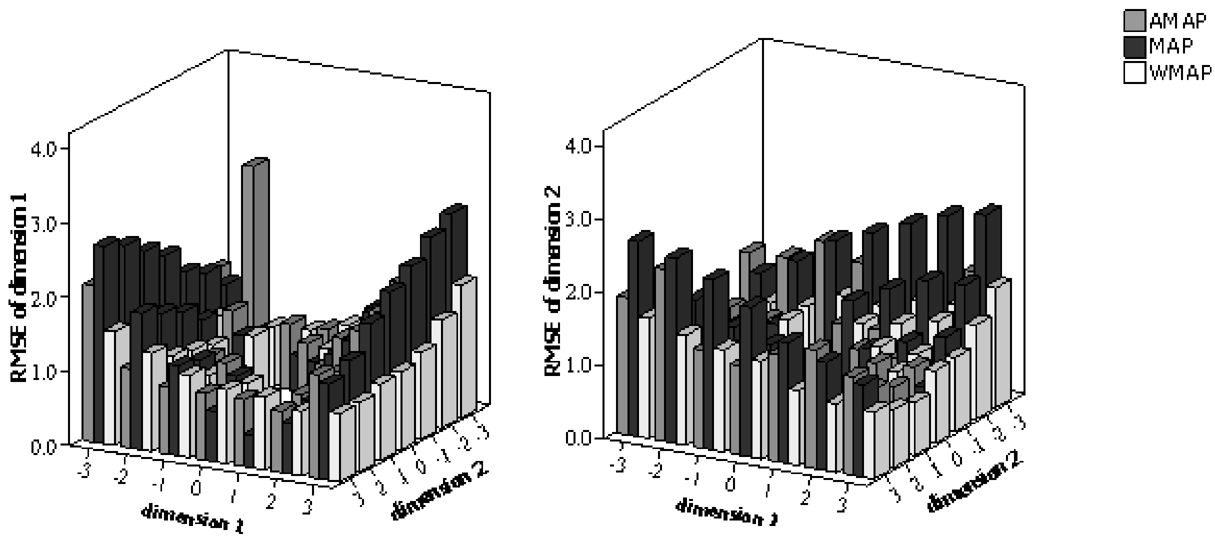


Figure 4 shows the conditional RMSE of MCATs with 10, 30, and 50 items when the θ combination were not correspondent with the prior distribution. In Figure 4a, when the test length was 10 items, the further the θ combination was away from the centroid of the prior distribution, the higher the RMSE of MAP (dark bars). The RMSE of AMAP (gray bars) and WMAP (white bars) were lower than that of MAP. These results indicate that AMAP and WMAP were less influenced by the unsuitable choice of the prior distribution. However, the RMSE of AMAP were quite high at some extreme θ levels, particularly at +3 and −3 logits in both dimensions.

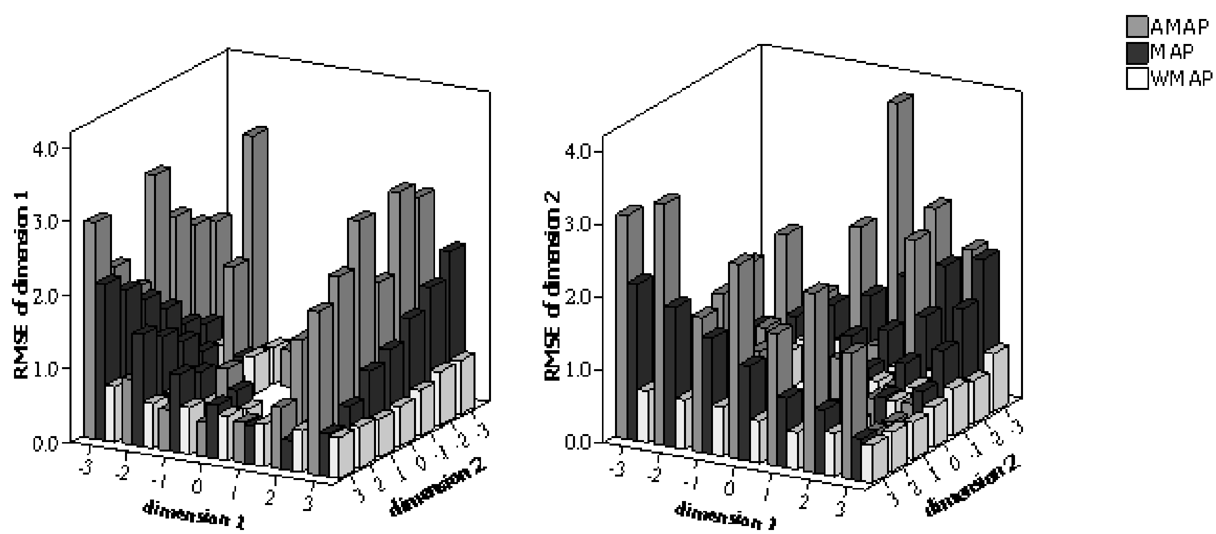
When the test length increased (Figures 4b and 4c), the RMSE of all three MAP estimators decreased. AMAP and WMAP were still more robust to the unsuitable choice of the prior distribution than the traditional MAP. However, the RMSE of AMAP increases to very high values at the extreme θ levels when the test length increased to 50 items. This result, which is similar to the results for bias, indicates that using AMAP will decrease the precision of θ estimation at the extreme θ levels when the prior distribution is not suitable.

Figure 4. The Conditional RMSE When the θ Combinations Were Not Correspondent With the Prior Distribution

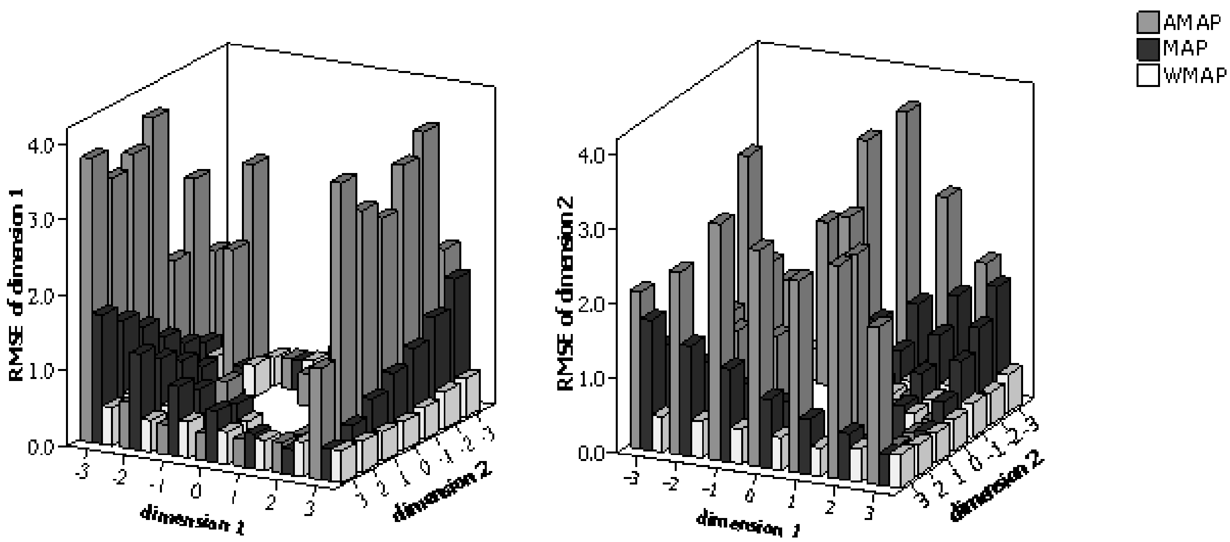
a. 10 Items



b. 30 Items



c. 50 Items



Summary. When the θ combinations were correspondent with the prior distribution, MAP showed less RMSE than AMAP and WMA. When the test length increased, the RMSE of AMAP and WMAP were close to the results of MAP. When the θ combinations were not correspondent with the prior distribution, the RMSE of MAP was higher than AMAP and WMAP. AMAP and WMAP estimations were robust to the unsuitable choice of the prior distribution. WMAP was much more stable than AMAP, especially at the extreme θ levels.

Discussion and Conclusions

This research compared the precision of θ estimation of two modified MAP methods and traditional MAP estimation in MCAT. Results indicated that when the prior distribution is appropriate, these two modified MAP methods can effectively reduce the regression bias of the traditional MAP method; however, they somewhat increase the RMSE when the test length is short (10 items). When the test length increases to 30 or 50 items, the RMSE of the two modified MAP estimators were close to the results of the traditional MAP estimation, and regression bias was eliminated.

AMAP estimation reduced the regression bias by changing the prior distribution after every item was administered. It moves the entire prior distribution toward the temporary θ estimates and treats it as the new prior distribution. This means that every temporary θ estimate has its own prior distribution. This method is similar to maximum likelihood estimation but it allows perfect and zero scores to be estimated by using the new prior distribution. However, if only the prior mean vector is changed but not the variance-covariance matrix,as was done in this study, the estimators will not be influenced by the original prior distribution and become very inaccurate at the extreme θ levels. The longer the test, the less the estimates are influenced by the original prior distribution. This might be the reason why there were higher bias and RMSE at the extreme θ levels when using AMAP estimation. Further research is needed to evaluate the effects of both the prior mean vector and the variance-covariance matrix are taken into consideration when using AMAP estimation.

WMAP estimation reduces the regression bias by adjusting the degree of use of the prior distribution. If the estimator is close to the centroid of the prior distribution, the weighted value is close to 1 and takes the complete prior distribution into account in the estimation process. If the estimator is away from the centroid of the prior distribution, it means that the possibility of the estimator coming from the prior distribution is low and the prior distribution is less suitable to be used. Hence, the further the estimator is away from the centroid of the prior distribution, the lower the weighted value of the prior distribution. This is the reason why WMAP estimation can reduce the regression bias and is quite robust to the unsuitable choice of the prior distribution. However, in short MCATs, the cost of using WMAP estimation is reducing the usage of the prior distribution and higher RMSE when the estimator is close to the prior distribution. Fortunately, when the test length increased, the influence of the prior distribution was relatively lower than the likelihood function, and the RMSE of the WMAP estimation was close to the result of the traditional MAP. It means that WMAP estimation can reduce the regression bias and preserve the precision of θ estimation.

These two modified MAP estimation were more robust to the unsuitable choice of the prior distribution than traditional MAP estimation in MCAT, and WMAP estimation appeared to be much more stable that AMAP estimation. The results of this research suggest that if the test administrator is highly confident of the prior distribution and is not concerned about regression bias, then using the traditional MAP estimation in MCAT is fine. If not, WMAP estimation can be used to reduce the regression bias and the influence of the unsuitable choice of the prior distribution. The WMAP is really adaptive to “all” the examinees.

References

Adams, R. J., Wilson, M. R., & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, *21,* 1-23.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561-573.

Bock, R. D. & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measuremen*t, *6,* 431-444.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika*, *46*, 443-459.

Chen, P.-H. (2006). The influences of the estimation methods on the precision of ability estimation in multidimensional computerized adaptive testing. *Educational Psychology Bulletin*. 38(2), 193-210. (in Chinese)

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, 39*, (Series B), 1-38.

Fischer, G. H. (1973). The linear logistic model as an instrument in educational research. *Acta Psychologica, 37*, 359-374.

Fischer, G. H., & Pononcy, I. (1994). An extension of the partial credit model with an application to the measurement of change. *Psychometrika, 59*, 177-192.

Johnson, D. E. (1998). *Applied multivariate methods for data analysts.* CA: Brooks/Cole Publishing.

Linacre, J. M. (1989). *Many-faceted Rasch measurement.* Chicago: MESA Press.

Luecht, R. M. (1996). Multidimensional computerized adaptive testing in a certification or licensure context. *Applied Psychological Measurement*, *20,* 389-404.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science, Calcutta*, *12*, 49-55.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Raiche, G., Blais, J.-G., & Magis, D. (2007). Adaptive estimators of trait level in adaptive testing: Some proposals. Paper presented in the 2007 GMAC Conference on Computerized Adaptive Testing. Minneapolis.

Rasch, G. (1960). *Probabilistic models for some intelligent and attainment tests.* Copenhagen: Institute of Educational Research. (Expanded edition, 1980. Chicago: The University of Chicago Press.)

Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika*, *61*, 331-354.

Wang, W.-C., & Chen, P.-H. (2004). Computerized adaptive testing for multidimensional between- and within-items. *Applied Psychological Measurement*, *28*, 1-22.

Weiss, D. J. & Mcbride, J. R. (1984). Bias and information of Bayesian adaptive testing. *Applied Psychological Measurement, 8,* 273-285.

Wilson, M. R. (1992). The partial order model: An extension of the partial credit model. *Applied Psychological Measurement*, *16,* 309-325.

Wu, M., Adams, R. J., & Wilson. M. R. (1998). *ACER ConQuest: Generalized item response modeling software*. Camberwell, Victoria: Australian Council for Educational Research.